João Baptista Cardia Neto

Reconhecimento de Faces 3D com Kinect

Bauru

December 2014

João Baptista Cardia Neto

Reconhecimento de Faces 3D com Kinect

Dissertation presented to the Computer Science Graduate Program, area of Applied Computing, of São Paulo State University (UNESP), as a requirement to obtain the title of Master Degree in Computer Science.

Orientador: Prof. Dr. Aparecido Nilceu Marana

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Graduate Program in Computer Science

Bauru

December 2014

Cardia Neto, João Baptista.
Reconhecimento de faces 3D com Kinect / João Baptista Cardia Neto -- São José do Rio Preto, 2014
65 f. : il., tabs.

Orientador: Aparecido Nilceu Marana
Dissertação (mestrado) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Computação. 2. Processamento de imagens - Técnicas digitais. 3. Reconhecimento facial (Computação) 4. Biometria. 5. Imagem tridimensional. 6. Sistemas imageadores. I. Marana, Aparecido Nilceu. II. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 518.72:76

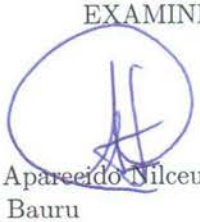
Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

João Baptista Cardia Neto

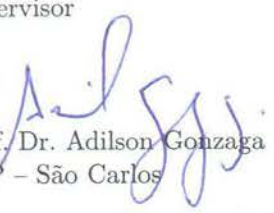
3D Face Recognition Using Kinect

Dissertation presented to the Computer Science Graduate Program, area of Applied Computing, of São Paulo State University (UNESP), as a requirement to obtain the title of Master Degree in Computer Science.

EXAMINERS



Prof. Dr. Aparecido Nilceu Marana
UNESP - Bauru
Supervisor



Prof. Dr. Adilson Gonzaga
USP – São Carlos

Prof. Dr. Antonio Carlos Sementille
UNESP - Bauru

Bauru
December 2014

Agradecimentos

Agradeço principalmente aos meus pais por terem serem me amparado e não me abandonado em momento algum, mesmo estando em um dos momentos difíceis da minha vida. Também ao Prof. Nilceu por sua paciência e orientação presente, sem ele não seria possível desenvolver o trabalho, muito menos chegar nos resultados que conseguimos.

Também gostaria de agradecer a CAPES pelo apoio financeiro. Por fim deixo também meu agradecimento ao PPGCC e a UNESP pela oportunidade de desenvolver esse projeto de pesquisa.

Resumo

Para identificação de pessoas o reconhecimento fácil possui várias vantagens sobre outros tipos de biometria, principalmente por sua alta universalidade, coletabilidade e aceitabilidade. Quando lidando com reconhecimento de faces 2D vários problemas aparecem, normalmente relacionados com pose, iluminação e expressão facial. Para aumentar a performance de métodos de reconhecimento facial vários algoritmos que utilizam modelos 3D foram propostos, uma vez que esse tipo de dado permite maior facilidade para tratamento dos problemas já mencionados. Com um modelo 3D é possível rotacionar a face em qualquer eixo, projetar iluminação e corrigir deformações ocasionadas por expressão. Os maiores problemas com reconhecimento de faces 3D estão vinculados com seus scanners, o alto custo e a forma intrusiva que eles funcionam. Alguns scanners 3D necessitam que a pessoa fique parada por todo o tempo de captura do modelo e, portanto, limitando a sua aplicação. Uma alternativa para scanners 3D tradicionais é a utilização do Kinect, um dispositivo criado pela Microsoft para aumentar a interação dos usuários com jogos no Xbox 360. O maior problema com o Kinect é que ele gera imagens de baixa resolução, dificultando a utilização desses dados para o reconhecimento de faces 3D. O principal objetivo dessa dissertação é analisar alguns métodos que foram propostos para o reconhecimento de faces 3D e propor novas formas de realizar essa função utilizando os dados do Kinect, com isso propomos um método que combina os descritores 3DLBP e o HAOG. Resultados experimentais obtidos no database EURECOM 3D mostram que a fusão dos métodos melhora o desempenho de ambos. Também foram propostas formas de melhorar a qualidade das faces quando houver obstrução parcial da face, usando preenchimento simétrico.

Palavras-chaves: *Kinect, reconhecimento de faces 3D, Biometria.*

Abstract

For person identification, facial recognition has several advantages over other biometric traits due mostly to its high universality, collectability, and acceptability. When dealing with 2D face images several problems arise related to pose, illumination, and facial expression. To increase the performance of facial recognition, 3D methods have been proposed and developed, since working with 3D objects allow us to handle better the aforementioned problems. With 3D objects, it is possible to rotate the face around any axis, generate illumination that matches the one in the environment and even correct the deformation in the model due to facial expression. The main problems with 3D facial recognition are: the high cost of 3D cameras that have been generally employed, and the intrusive way that such devices work. Some of them require that the subject remain completely still for several minutes while scanning, limiting, therefore, the application deployment for uncontrollable environments. One alternative to those expensive cameras is the Kinect, a device developed by Microsoft to enhance gaming in the Xbox 360 console. Due to its capacities to generate depth images, Kinect is a candidate device to be used for 3D face recognition, replacing the traditional 3D cameras. The main problem with Kinect is that it generates low-resolution images, making difficult the task of precise facial recognition. The main objective of this dissertation was to assess some methods that have been proposed recently for 3D face recognition and to propose new methods for this task utilizing the data generated by Kinect devices. Then, we have proposed a new method that combines 3DLBP and HAOG features. Experimental results obtained on the EURECOM 3D face database show that when 3DLBP and HAOG features are combined the results can be better than they are used alone. We have also proposed a method that increases the facial recognition performance when the faces present partial obstructions, by utilizing a symmetric filling approach.

Key-words: *Kinect, 3D face recognition, Biometrics.*

Contents

1 Introduction **11**

1.1 Objectives . . . 12
1.2 Motivation . . . 12
1.3 Dissertation Structure . . . 12

2 Biometric Identification **14**

2.1 Personal Identification . . . 14
2.2 Biometrics Characteristics . . . 15
2.3 Biometric Systems . . . 17
2.4 Conclusion . . . 21

3 Related Work **22**

3.1 Introduction . . . 22
3.2 The Kinect Sensor . . . 23
3.3 3D Face Recognition Methods . . . 24
3.3.1 3D Face Recognition with Keypoint Detection and Local Features . . . 24
3.3.1.1 Keypoint Detection . . . 24
3.3.1.2 3D Feature Extraction . . . 25
3.3.1.3 2D Feature Extraction . . . 25
3.3.1.4 Matching . . . 26
3.3.2 Face Recognition under Varying Pose, Expression, Illumination and Disguise utilizing Kinect . . . 27
3.3.2.1 3D Face Cropping and Pose Correction . . . 27
3.3.2.2 Symmetric Filling . . . 27
3.3.2.3 Multi-Modal Sparse Coding . . . 27
3.3.3 Face Recognition Based on Local Low-Level Features . . . 28
3.3.3.1 Feature Extraction . . . 28
3.3.3.2 Support Vector Machines . . . 30
3.3.3.3 Feature-Level Fusion . . . 32
3.3.3.4 Score-Level Fusion . . . 33
3.3.4 RGB-D Face Recognition utilizing Kinect . . . 34
3.3.4.1 Extraction of Entropy and Visual Salience Maps . . . 34
3.3.4.2 Histrogram of Oriented Gradient . . . 35
3.3.5 Comparison of the 3D Face Recognition Methods . . . 35
3.4 Local Binary Pattern . . . 38
3.4.1 3D Local Binary Pattern . . . 39
3.4.2 Gradient-LBP . . . 40
3.5 Histogram of Averaged Oriented Gradients . . . 41
3.6 Conclusion . . . 42

4 Proposed Method **44**
4.1 Face Normalization 44
4.2 Feature extraction 44
4.3 Fusion Strategy 44
4.4 Generation of Depth Maps from the Cloud Points 46

5 Experiments and Results **48**
5.1 EURECOM Kinect Dataset 48
5.2 Experiments Protocol and Performance Measurement 49
5.3 Experiments 50
5.3.1 Experiment 1 50
5.3.2 Experiment 2 54
5.3.3 Experiment 3 56

6 Conclusion **58**
6.1 Contributions 59
6.2 Future work 59

Bibliography **60**

List of Figures

Figure 1 Different types of authentication that was utilized in 1994. It is important to note that concern with secure ways of assuring an individual identity is not new (MILLER, 1994). 15

Figure 2 Examples of biometrics characteristics: a) ear, b) face, c)facial thermogram, d) hand thermogram, e)hand vein, f)hand geometry, g) fingerprint, h)iris, i)retina, j)signature, and k) voice. The examples are taken from (JAIN; MALTONI, 2003). 16

Figure 3 Enrollment, verification and identification stages of a biometric system (PRABHAKAR; PANKANTI; JAIN, 2003). 18

Figure 4 Biometric system error rates (PRABHAKAR; PANKANTI; JAIN, 2003). 20

Figure 5 ROC Curve, taken from (JAIN; ROSS; NANDAKUMAR, 2011). 20

Figure 6 The Kinect sensor. 23

Figure 7 Binary mask cropping the areas from the face (LEI; BENNAMOUN; EL-SALLAM, 2013). 29

Figure 8 An example of the four geometric features (LEI; BENNAMOUN; EL-SALLAM, 2013). 30

Figure 9 Comparison of histograms generated from the same subject (LEI; BENNAMOUN; EL-SALLAM, 2013). 31

Figure 10 CMC curves comparing the identification results (LEI; BENNAMOUN; EL-SALLAM, 2013). 36

Figure 11 ROC curves comparing the verification results (LEI; BENNAMOUN; EL-SALLAM, 2013). 37

Figure 12 Example of how to calculate the LBP operator. 39

Figure 13 Example of a $LBP_{(4,1)}$ neigborhood of a pixel. The black dots are the sampling points in the red circle. 39

Figure 14 The full process of the 3DLBP proposed by (HUANG; WANG; TAN, 2006). Each of the differences is encoded into the layers (layer 2, 3 and 4) and the signal into the layer 1. 40

Figure 15 Diagram of the proposed method for 3D face recognition using Kinect data. The proposed method combines the 3DLBP and HAOG methods in order to provide a better performance. 45

Figure 16 Two types of depth maps utilized in the proposed work. a) Depth map generated directly by the Kinect device. b) Depth map generated from the cloud points outputted by the Kinect device after the symmetric filling and approximation processes. 46

Figure 17 The generation of new depth maps based on the cloud points 47

Figure 18 A set of images from a subject in the EURECOM database, in which it is possible to see different poses and facial expressions in the RGB (top row) and the depth map (bottom row) images. 49
Figure 19 An example of annotated face landmarks in the EURECOM database: left and right eyes, the tip of the nose, left and right side of the mouth and the chin. 49
Figure 20 CMC curves obtained for the 3DLBP, HAOG, 3DLBP fused with HAOG, Salience and Entropy map, FPLBP, HOG, 3DPCA. 51
Figure 21 ROC curves obtained for the 3DLBP, HAOG, and 3DLBP fused with HAOG, for the depth map generated directly by Kinect and the depth map obtained from the Kinect cloud points, using the Set 1. 52
Figure 22 CMC Curve comparing different values for w_1 and w_2, this fusion was made utilizing only the new depth maps. 53
Figure 23 CMC Curve comparing different values for w_1 and w_2, this fusion was made utilizing only the original Kinect depth maps. 54
Figure 24 ROC curves for the 3DLBP, HAOG, and the fusion between them. The results were obtained utilizing the Set 2 (which includes occlusion). 55
Figure 25 CMC Curve comparing different values for w_1 and w_2. This experiments have an image with occluded face in the probe. 56

List of Tables

Table 2 Comparison among biometric characteristics (JAIN; ROSS; PRABHAKAR, 2004). 17

Table 3 Comparison among different 3D scanners (LI et al., 2013). 24

Table 4 Comparison between methods of 3D face recognition (LEI; BENNAMOUN; ELSALLAM, 2013). 36

Table 5 Results for recognition tests in the database CurtinFaces with the method proposed by (LI et al., 2013). In this test there are pose X facial expression variations. 37

Table 6 Results for recognition tests in the database CurtinFaces with the method proposed by (LI et al., 2013). In this test there are illumination X facial expression variations. 38

Table 7 Recognition error rates for the experiment utilizing the ACDNP features on the face images from the Set 1. 57

List of Acronyms

3DLBP	*3D Local Binary Pattern*
CMC	*Cumulative Match Characteristic*
DCS	*Discriminant Color Space*
EER	*Equal Error Rate*
FAR	*False Accept Rate*
FMR	*False Match Rate*
FNMR	*False Non-Match Rate*
FRGC	*Face Recognition Grand Challenge*
FRR	*False Rejection Rate*
FTC	*Fail To Capture*
FTE	*Fail To Enroll*
GAR	*Genuine Accept Rate*
HAOG	*Histogram of Averaged Oriented Gradients*
HOG	*Histogram of Oriented Gradients*
ICP	*Iterative Closest Point*
LBP	*Local Binary Pattern*
PCA	*Principal Component Analysis*
PIN	*Personal Identification Number*
RDF	*Random Decision Forest*
ROC	*Receiver Operating Characteristics*
SIFT	*Scale Invariant Feature Transform*
SRC	*Sparse Representation Classifier*
SVM	*Support Vector Machine*

Chapter 1

Introduction

Daily, thousands of individuals need to have their identity assured. Assuring you are the person you claim to be is not a trivial task. There are many ways a subject can use to ensure his identity: a PIN (personal identification number), a password, an ID card, a social security number, a key, a document just to name a few. Such identification approaches are based on "something you know" (PIN, Password) or "something you possess" (ID card, key, document) (BOLLE; PANKANTI, 1998). The main problem with these approaches is the possibility of losing your possessions, forgetting your knowledge or having them acquired by an impostor. Mastercard company, for instance, estimates that $450 million per year is spend with fraudulent credit cards. In LexisNexis (2013) it is shown the increase of cost from fraudulent transactions, specially from the online channel. The same study links this increase with the easiness of obtaining credit card numbers and anonymity. Besides online channel in-person fraud increased from 2012, going from 58% to 62%.

These problems shows that the traditional identification methods cannot guarantee if the person being identified is genuinely the correct individual or is an impostor. It becomes obvious that biometrics based methods help to attenuate these problems since they utilize physical (face, fingerprint, or iris) or behavioral characteristics (gait, signature, or voice) for individuals identification. Biometrics can be effective measure to reduce the amount of fraudulent transactions (BOLLE; PANKANTI, 1998).

Human beings utilize faces naturally to recognize different individuals and in automatic identification systems, the use of facial traits presents several advantages over other types of biometric characteristic. For instance, is possible to capture a face from a considerable distance, in a covert manner, this makes face one of the best choices when discretion is a requirement of the identification application.

In constrained applications automated facial recognition systems proposed so far can perform better than the human visual system (LI; JAIN, 2011), but the real challenge arise when dealing with uncontrolled environments, where the faces can present severe variations in illumination, pose, and facial expressions.

One way to increase the automated facial recognition system's accuracy is to utilize

different sensing devices such as 3D or infrared cameras (KAKADIARIS et al., 2007). However, the main issue with these devices is their high costs.

1.1 Objectives

The main objective of this dissertation was to assess some methods that have been proposed recently for 3D face recognition and to propose new methods for this task utilizing the data generated by Kinect v1 devices. The main problem with Kinect v1 is that it generates low-resolution images, making difficult the task of precise facial recognition. On the other hand, Kinect can be an alternative to the high cost cameras that have been used for 3D face recognition. Thus, this work aims to assess the advantages and disadvantages of using the Microsoft Kinect based technology as a cheaper way to carry out 3D face recognition for use in biometrics identification applications.

1.2 Motivation

One of the main advantages of utilizing face recognition over other types of biometrics traits is the easiness to obtain a biometric sample from an individual. With the current technology, it is possible, for instance, to obtain images of an individual's face at a distance, in a discrete manner, allowing a range of new types of identification applications.

The main problem with 2D face images is that illumination and pose variations can dramatically decrease the identification performance. Since the face is a 3D object one of the best way to deal with the difficulties impose by pose and illumination is utilizing 3D face model.

The main problem with 3D face recognition methods is the high cost of the cameras that have been proposed for 3D face recognition, and the way that these cameras have to be used in order to capture properly the 3D face information. Depending on the type of camera the individual need to be completely still with the eyes closed while his/her face is being scanned. This is unfeasible in applications that need some sort of secrecy, or that cannot rely on a high degree of user cooperation.

Kinect device can be an interesting alternative to the expensive cameras for the 3D face recognition. The system that utilize the Kinect can operate it in the same way the systems that use conventional cameras, but diminishing the need for user interaction. Another advantage of using Kinect is the possibility to retrieve simultaneously RGB and depth information from the same scene.

The relevance of this work becomes evident, since one way to increase the performance of facial recognition systems is utilizing data extracted from depth information, and Kinect device seem to be ideal to get this kind of data.

1.3 Dissertation Structure

Besides this introductory chapter, this dissertation is divided into six other chapters.

The chapter 2 refers to Biometrics, discussing its main concepts addressed in this work.

The chapter 3 presents methods for 3D face recognition found in literature, focusing in the methods with greater performance. A method that extracts local features from a face, and was utilized as a starting point for our work, is presented in the end of this chapter.

The chapter 4 presents the proposed method: the fusion between the 3DLBP and HAOG operators.

The chapter 5 presents the material and the methodology of our work.

The chapter 6 shows the results that were obtained in our work.

The chapter 7 discusses the results obtained in this work and presents its conclusions.

Chapter 2

Biometric Identification

This chapter presents the definition of biometric identification, its benefits over the other types of assuring a person identity, how a biometric system works and the different types of biometric characteristics. Before discussing what biometric identification is and how it can make a system robust to fraud, it is imperative to understand these concepts and the flaws of traditional methods of person identification. With this is possible to discuss the importance of researching and building different biometric system.

2.1 Personal Identification

Personal identification is the act of linking a person with an identity. It can be categorized in verification and identification (BOLLE; PANKANTI, 1998).

Verification deals with denying or confirming an individual claimed identity, whilst identification deals with establishing an individual identity. When utilizing a set of already known identities to perform this task it can be called a closed identification problem, otherwise it is an open identification problem (BOLLE; PANKANTI, 1998).

One way of doing this is utilizing some sort of object (e.g. document) or knowledge that a person is supposed to possess (PRABHAKAR; PANKANTI; JAIN, 2003). This is illustrated by Miller (1994), as shown in the Figure 1. The identification based on what you know or what you have has several problems. The stolen of an ID card can lead to a stranger being wrongly identified as the true owner of that object and, consequently, gaining access to private information or area. This illustrates that the traditional form of person identification cannot be utilized to verify if the carrier of an identification document is really its owner or an impostor. Even when combining something you have (a card, for instance) with something you know (a password, for instance) the human component is fragile to flaws. According to Hayday (2002) 81% of the Internet users have a weak (common) password, such as the word password or their date of birth, and 30% write their password in files. Another common security flaw is to write the PIN on the back of the credit card.

When utilizing biometric characteristics it is possible to attenuate problems like the

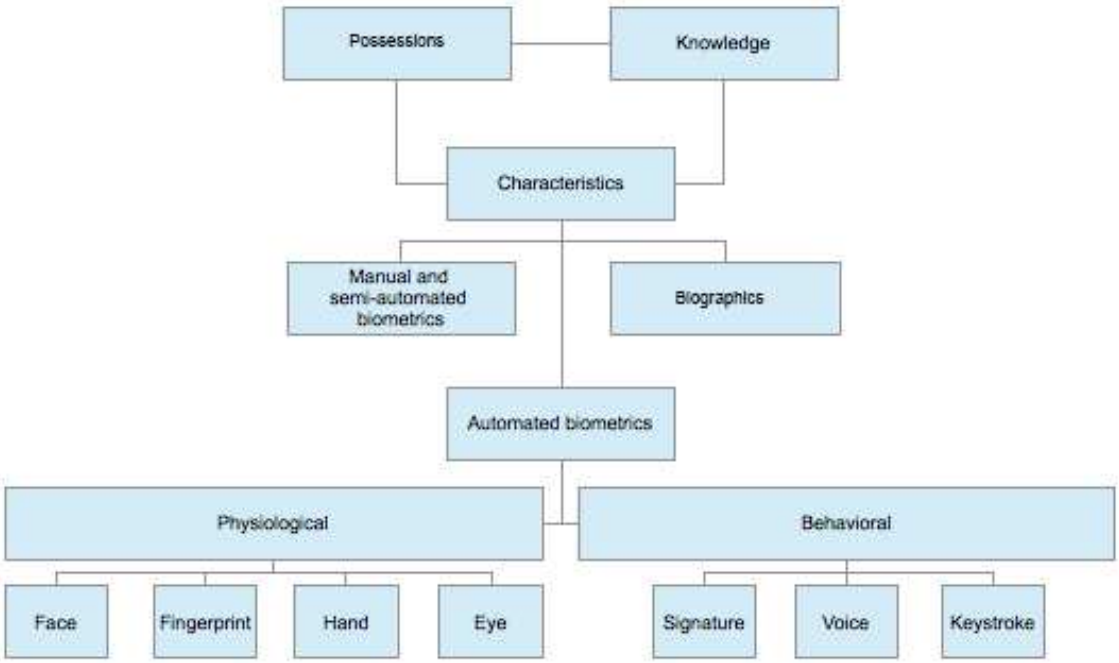


Figure 1 – Different types of authentication that was utilized in 1994. It is important to note that concern with secure ways of assuring an individual identity is not new (MILLER, 1994).

aforementioned ones. However, this does not mean that a biometric system is completely fraud free. For instance, if a system is utilizing fingerprint for person identification and does not have any sort of treatment to avoid fraud then it is possible to use artificial finger models from a genuine individual to fool the system. In a case reported by Estado de São Paulo newspaper (ESTADÃO, 2012) the owner of one driver school charged students to fake their presence utilizing fingerprint silicon molds.

2.2 Biometrics Characteristics

Biometrics is the automatic recognition of an individual identity based on its characteristics. For this task, physical/physiological or behavioral traits can be considered (PRABHAKAR; PANKANTI; JAIN, 2003). Physical/physiological characteristics deals with the anatomical or individual living functions (e.g. fingerprint of facial thermogram). Behavioral characteristics are those traits that can identify an individual through a particular way of doing some kind of task (e.g. gait, and signature). In the Figure 2 some examples of physical/physiological and behavioral characteristics are shown.

It is possible to use any human characteristic for biometric recognition since it satisfies a few requirements. Jain & Maltoni (2003) defines them as:

- Universality: Most of people need to posses that characteristic;
- Distinctiveness: That characteristics need to be distinct for distinct persons;
- Permanence: That characteristics should not change over time;

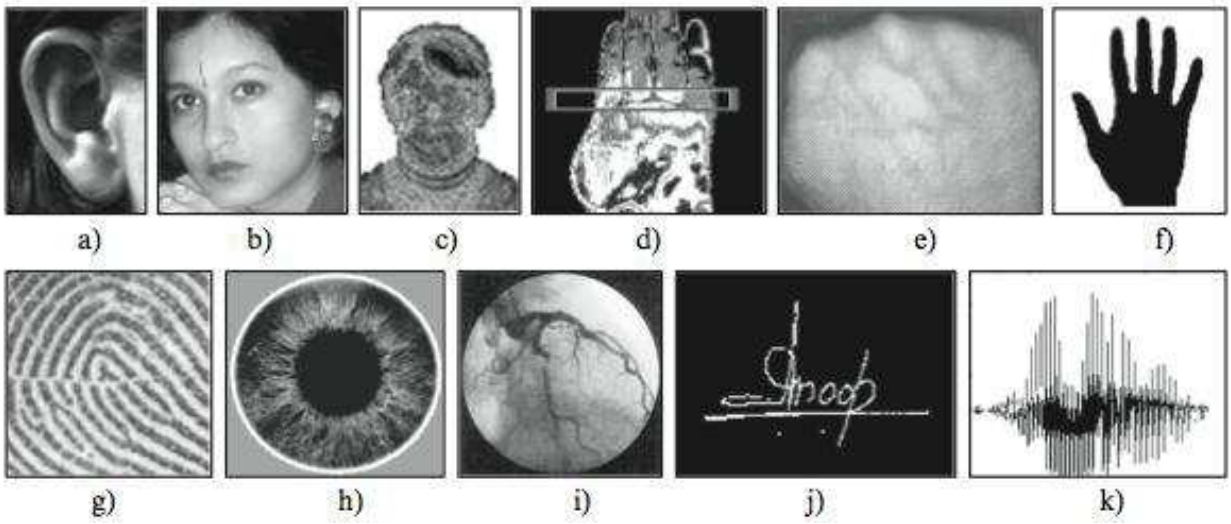


Figure 2 – Examples of biometrics characteristics: a) ear, b) face, c)facial thermogram, d) hand thermogram, e)hand vein, f)hand geometry, g) fingerprint, h)iris, i)retina, j)signature, and k) voice. The examples are taken from (JAIN; MALTONI, 2003).

- Collectability: That characteristics must be quantitatively measured;
- Performance: That characteristic must provide high accuracy and robustness, while providing low processing time and computational costs;
- Acceptability: That characteristic must be accepted culturally and socially by the people to be identified;
- Circumvention: That characteristic must be difficult to circumvent.

The last three requirements are more relevant to practical biometric systems. The Table 2 shows a comparison among some biometric characteristics in respect to these requirements (JAIN; ROSS; PRABHAKAR, 2004). It is important to emphasize that there is no perfect biometric. When deciding what biometric trait to utilize it is essential to think about the environment constraints and the characteristics of the people group to be identified. Due to its characteristic, face would be a proper choice in an application that needs to capture the biometric sample in a covert manner, the same cannot be said for fingerprint. This does not mean that one characteristic is superior to other, but just that it works better on some specific application. Even though concerning about security is important, other issues should be taken into consideration, such as reasonable resource requirements, harmless to the users, accepted by the intended population, privacy, and robustness to fraud (JAIN; MALTONI, 2003).

Analyzing the Table 2 one can conclude that behavioral characteristics are more susceptible to fraud than the physiological ones because it is easier to mimic someone else's behavior than its physiological traits.

Another important thing to notice is that face has higher universality than fingerprint, higher collectability than iris, and higher acceptability than fingerprint and iris. In other hand,

Table 2 – Comparison among biometric characteristics (JAIN; ROSS; PRABHAKAR, 2004).

Biometric	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	High	Low	Medium	High	Low	High	Low
Fingerprint	Medium	High	High	Medium	High	Medium	Medium
Hand Geometry	Medium	Medium	Medium	High	Medium	Medium	Medium
Iris	High	High	High	Medium	High	Low	High
Hand Vein	Medium	Medium	Medium	Medium	Medium	Medium	High
Ear	Medium	Medium	High	Medium	Medium	High	Medium
Keystroke	Medium	Medium	Low	Medium	Low	Medium	Medium
Odor	High	High	High	Low	Low	Medium	High
DNA	High	High	High	Low	High	Low	High
Facial Thermogram	High	High	Low	High	Medium	High	High
Retina	High	High	Medium	Low	High	Low	High
Signature	Low	Low	Low	High	Low	High	Low
Voice	Medium	Low	Low	Medium	Low	High	Low
Gait	Medium	Low	Low	High	Low	High	Medium

face is more susceptibility to fraud, has lower performance, permanence and distinctiveness than the aforementioned characteristics.

Since it is easy to capture high quality images with current technology face stands out when discretion or little user interaction is required.

2.3 Biometric Systems

A pattern recognition system that utilizes a feature vector based on any biometric characteristic to assure an individual identity is a biometric system. A Biometric system will normally operates in one of two modes: Identification or Verification (PRABHAKAR; PANKANTI; JAIN, 2003).

Verification mode is when the user claims to be a certain person and the system compares the biometric sample probed with his template stored in the database. With this is possible to answer the question "Is this person who she or he claims to be?". Identification mode is when given a biometric sample the system searches through all the templates stored in the database trying to find a match. With this is possible to answer the question "Who is this person?". These two distinct modes are typically used for positive or negative person recognition, respectively (PRABHAKAR; PANKANTI; JAIN, 2003).

Before recognizing an individual, it is necessary to enroll him/her in the gallery. The enrollment starts with the subject providing a biometric sample and the system extracting and compressing that data into a template. It is important to assure that the captured sample has an acceptable quality, the system checks for it and, if necessary, asks for another sample. The template can be stored in a central database or in a type of removable media (e.g. flash drive).

In verification mode, the user provides the biometric sample and claims to be an individual. The system then recovers the template from the claimed user and compares both samples. If

there is a match, the system validates the individual identity. The method for feature extraction has to be the same as in the enrollment stage.

In identification mode, the individual only needs to provide his/her biometric, the system will be responsible for finding the associated identity. The biometric is processed and compared with all the templates in the database. The system will indicate an identity that is most similar to the sample or will indicate that there is no such individual in the database. The Figure 3 exemplifies these processes.

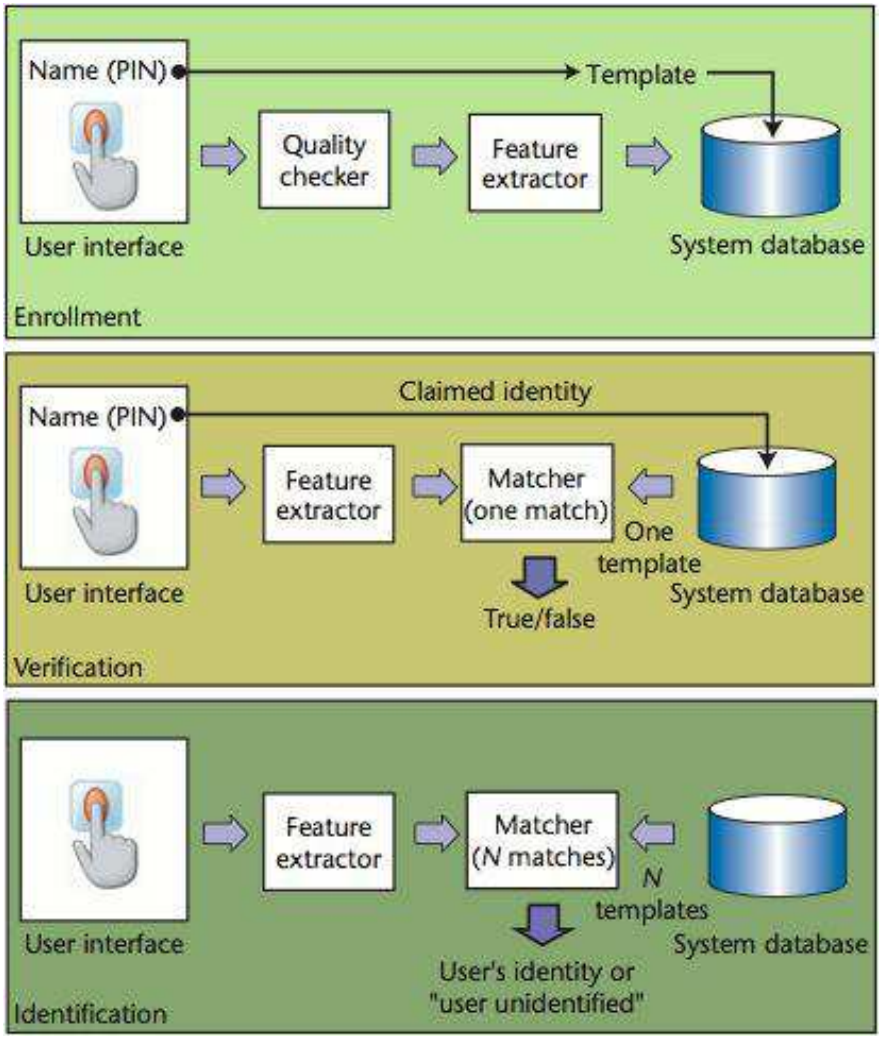


Figure 3 – Enrollment, verification and identification stages of a biometric system (PRABHAKAR; PANKANTI; JAIN, 2003).

A biometric system can be classified in one of the seven categories defined by Wayman (2002):

- Cooperative versus Non-cooperative: Does the user wish to be identified?
- Overt versus Covert: Does the user knows that he is being identified?
- Habituated versus Non-habituated: Does the user often submits to the identification?
- Attended versus Non-attended: There is a human operator helping the system?

- Standard environment: What is the environment that the system will operate?
- Public versus private: The users are employees (private) or clients (public)?
- Open versus closed: Does the system needs to utilize standards to provide interoperability between systems?

Due to noise, environmental conditions, changes in the person's traits, and even how the user interacts with the sensor, two samples of a biometric characteristic from the same person will never be the same. Therefore, the system will have a matching score and this score will determine if the comparison succeeds or fails. To do so a threshold t must be set to regulate the system decision. If the score is greater or equal to t then the pair of characteristics belong to the same individual. Otherwise they do not.

Due to their nature, biometric systems can present two types of errors (PRABHAKAR; PANKANTI; JAIN, 2003):

- False match: A biometric from two different persons are considered to come from the same person;
- False non-match: A genuine biometric comparison is taken as an impostor.

To work properly a biometric system must make a trade-off with these two error rates. If the system designer decides to decrease the FMR (false match rate) then the FNMR (false non-match rate) will increase. On the other hand, if the designer tries to facilitate the user login, the FMR will increase. The Figure 4 shows the correlation between the two errors rates.

Other error rates are the FAR (False Accept Rate) and the FRR (False Rejection Rate), they are analogous with the FMR and FNMR. The FAR is the rate in which a false subject is accepted as genuine and the FRR is the rate in which a genuine match is categorized as an impostor.

Given both, the FAR and FRR, it is possible do calculate the Equal Error Rate (EER). The EER is a security level measurement that identifies a threshold when the FAR and FRR have the same value.

While FMR and FNMR are intrinsic system errors, there are others that can happen due to conditions that cannot be controlled: the fail to capture (FTC) and fail to enroll (FTE) (PRABHAKAR; PANKANTI; JAIN, 2003).

Since a biometric system has to make a tradeoff between two errors rates (FNMR X FMR) it is not possible to assess its performance with a single number, for evaluating this kind of system, a performance curve is necessary (MARTIN et al., 1997). The Receiver Operating Characteristics (ROC) curve is one way to understand the performance of one biometric system. A ROC curve is a two-dimensional graph where the true positive rate is plotted on the Y axis while the false positive rate is plotted on the X axis (FAWCETT, 2006). In a ROC curve for evaluating biometric system usually the Genuine Accept Rate (GAR) is plotted against the

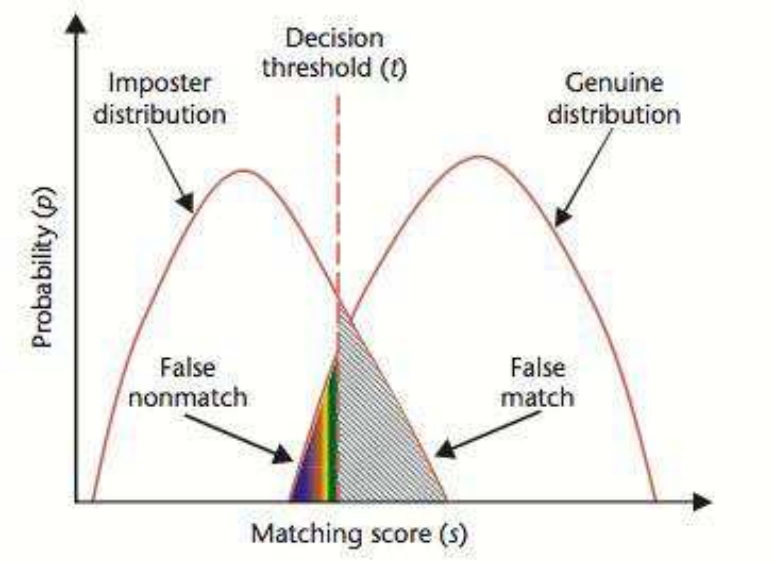


Figure 4 – Biometric system error rates (PRABHAKAR; PANKANTI; JAIN, 2003).

FAR (JAIN; ROSS; NANDAKUMAR, 2011). The Figure 5 shows a ROC curve for various thresholds. The GAR is defined as the fraction of genuine scores that exceed a certain threshold (JAIN; ROSS; NANDAKUMAR, 2011).

For forensic applications the FNMR is more important than the FMR, this is because normally this kind of application deals with criminal identification and it is very important not letting a suspect pass unidentified, even if it is needed to manually select the right identity from a group of matched subjects. For high-security applications occurs the opposite, a wrongfully identified subject cannot gain access to the system (PRABHAKAR; PANKANTI; JAIN, 2003).

For civilian applications a balance between the two rates is the desired scenario.

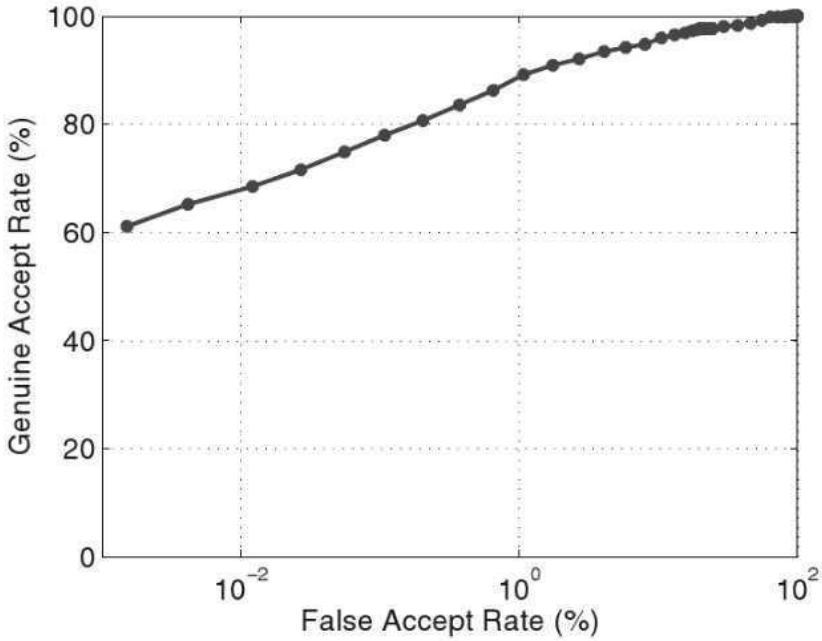


Figure 5 – ROC Curve, taken from (JAIN; ROSS; NANDAKUMAR, 2011).

2.4 Conclusion

In this chapter, the present definition of biometric was discussed and how it can benefit over other types of assuring a person identity. It was also discussed different types of error rates to assess performance for an identification method and how to represent the rates utilizing the ROC curve.

In the course of this chapter it became clear the importance of researching and building different biometric system and that there is no perfect biometric characteristic, there are different scenarios in which certain biometric can outperform the others.

In current days there is the necessity to make high security data available online, in a way that authorized personal can access it any time of the day. Since traditional ways of assuring a person identity cannot distinguish a genuine individual from an impostor, it is not feasible to utilize them.

Biometric systems can distinguish better impostor from genuine users. Then, it becomes a very promising candidate to substitute the legacy authentication and identification systems.

Chapter 3

Related Work

In this chapter, after a brief introduction and a concise description of Kinect, some related work are presented, starting with four methods for 3D facial recognition.

The first method, proposed by Mian, Bennamoun & Owens (2007), combines 2D and 3D data. For matching it employs feature based and holistic approaches as a way to surpass the deformation caused by facial expressions. The second method, proposed by Li et al. (2013), deals with low resolution 3D images, captured by Kinect devices, in very challenging environments. The third method, proposed by Lei, Bennamoun & El-Sallam (2013), is based on low-level features. The fourth method is also based on data captured by Kinect (GOSWAMI et al., 2013). It generates entropy maps for RGB and depth data, and visual saliency map for RGB data. This method utilizes HOG (Histogram of Oriented Gradients) features and a RDF (Random Decision Forest) classifier, in order to identify the subjects.

In the end of this chapter, a variation of the LBP (Local Binary Pattern) method, called 3DLBP (3D Local Binary Pattern), which deals with depth information for face recognition, and the face descriptor HAOG (Histogram of Averaged Oriented Gradients) are also presented.

3.1 Introduction

Utilizing face for recognizing another person is one of the most natural ways of people identification. However, even with the heavy research on automatic face recognition observed in the last two decades, there are few human identification applications in the real world based on such biometric trait. This is due to the fact that changes in illumination, pose, and facial expressions, besides faces obstructions, make the effective and robust 2D face recognition a very challenging task (LI; JAIN, 2011). One way to increase the automatic face recognition rates in real scenarios is to utilized the 3D information of the face.

One of the reasons that 3D recognition methods can overcome 2D face recognition is the possibility to correct pose and illumination. It is possible, for instance, to rotate or to project illumination in a 3D model in order to adjust changes, and sometimes it is unfeasible to do using only 2D data.

Another reason that 3D face recognition has gained popularity in the recent years is the advances in 3D scanning technology. Nevertheless, even with such advances, utilizing traditional 3D scanners are not feasible for real world applications because they are very expensive, very frail, and not very user friendly (LI; JAIN, 2011). These problems with traditional 3D scanners are discouraging since one of the main advantages of using the face as biometric characteristic is its acceptability and the possibility to capture the face image at a distance, in a covert way, that is, without any user collaboration. One alternative to reduce the problems regarding the traditional 3D scanners can be the use of Microsoft Kinect sensors.

3.2 The Kinect Sensor

The Kinect sensor is composed of a 3D depth sensor, a RGB camera, a microphone array, and a motorized tilt, as one can observe in the Figure 6. The 3D depth sensor is made of two components: an infrared emitter and an infrared camera. The depth data is calculated projecting an array of infrared dots at the scene and measuring the distortion caused by the rays reflected back to the camera (SHPUNT; ZALEVSKY, 2008). The sensor outputs a 320 X 240 depth grid with a resolution of 11 bits at 30 Hz (ZOLLHÖFER et al., 2011). In the official specification, the depth range has the value of 1.2-3.5 meters.

Figure 6 – The Kinect sensor.

The cameras have an angular field of view of 57 ° horizontally and 43 ° vertically (STOWERS; HAYES; BAINBRIDGE-SMITH, 2011). One important thing about Kinect is that the angle is relative to the horizon and not to the base of the device. The tilt engine possibility enables the device to scan the whole scene and find the better angle to capture the biometric sample.

In Table 3 (LI et al., 2013) one can see a comparison among different scanners. The columns list the time taken by the sensor to scan the individual, the time taken for charging (if needed), the size, the price, and the measurement accuracy of each scanner.

This comparison shows the importance of developing methods for facial recognition

utilizing Kinect, not just because of the price, but because the restrictions that other equipment have. Taking a closer look at the characteristics it is possible to observe that the Minolta takes 2.5 seconds to scan a person (and during this time the subject must remain completely still in front of the camera). This implicates that for extracting a biometric sample using this device, it would be necessary a great deal of user cooperation. With this restriction, some applications would be unfeasible. Even though the Kinect has a smaller precision, it can scan a person in a very small fraction of time and in a covert manner.

Table 3 – Comparison among different 3D scanners (LI et al., 2013).

Device	Speed (sec)	Charge Time	Size ($inch^3$)	Price (USD)	Acc. (mm)
3dMD	0.002	10 sec	N/A	>$50K	< 0.2
Minolta	2.5	no	1408	>$50K	0.1
Artec Eva	0.063	no	160.8	>$20K	0.5
3D3 HDI R1	1.3	no	N/A	>$10K	0.3
SwissRanger	0.02	no	17.53	>$5K	10
DAVID SLS	2.4	no	N/A	>$2K	0.5
Kinect	0.033	no	41.25	<$200	1.5-50

3.3 3D Face Recognition Methods

In this section, four methods proposed recently for 3D face recognition are presented, two of them are based on Kinect data.

3.3.1 3D Face Recognition with Keypoint Detection and Local Features

In their work Mian, Bennamoun & Owens (2007) proposed a multi-modal approach, which first detects and utilizes keypoints in the 3D model and extracts a local feature from its neighbor. From the 2D domain SIFT (Scale Invariant Feature Transform) features (LOWE, 2004) are extracted and, for the matching process, those features are fused. In this section, a brief overview of this method is presented.

3.3.1.1 Keypoint Detection

The keypoints are landmarks that can be identified with high repeatability in the same surface in the presence of noise and deformation (e.g. facial expression) (MIAN; BENNAMOUN; OWENS, 2007). The keypoint detection has as input a cloud point $F = [x_i, y_i, z_i]^T$ representing the face, the next step is to sample the face in uniform intervals. At each sample point p, a region is cropped from the face with radius r_1. The value of the radius is important because big values will make the cropped region sensible to noise and deformation, whilst very small values can hurt the descriptiveness of the feature. The mean vector m and the covariance C of the face L, in which $L_j = [x_j, y_j, z_j]^T$ are the cropped regions of the face and $j = 1, \cdots, n_l$ with n_l being the maximum number of points in the L_j region, can be calculated by:

$$m = \frac{1}{n_l} \sum_{j=1}^{n_l} L_j \tag{3.1}$$

$$C = \frac{1}{n_l} \sum_{j=1}^{n_l} (L_j L_j^T - mm^T) \tag{3.2}$$

Performing PCA (Principal Component Analysis) on the covariance matrix C results in the matrix of eigenvectors V:

$$CV = DV \tag{3.3}$$

being D the diagonal matrix of the eigenvectors of C.

It is possible to align the matrix L with its principal axes. This is done utilizing the Hotelling transform (GONZALEZ; WOODS, 2001):

$$L'_j = V(L_j - m)\{j = 1, \cdots, n_l\} \tag{3.4}$$

with L'_x and L'_y being the x and y components from the face represented by the cloud point L'. The next step for finding the keypoints is the calculation of δ.

$$\delta = \max(L'_x) - \min(L'_x) - (\max(L'_y) - \min(L'_y)). \tag{3.5}$$

If $\delta = 0$ then L' is planar or spherical. If $\delta \neq 0$ then L' has unsymmetrical variation in its depth. If $\delta \geq t_1$ then p is a keypoint. The values of t_1 and r_1 are defined empirically as 20mm and 2mm, respectively. Even if the parameters have small changes the algorithm will not suffer major differences, since these values are relative to the scale of the human faces.

3.3.1.2 3D Feature Extraction

For feature extraction Mian, Bennamoun & Owens (2007) use the extending tensor representation, presented on their previous works (MIAN; BENNAMOUN; OWENS, 2006) and (MIAN; BENNAMOUN; OWENS, 2006). In those works, local surface patches are quantified into three-dimensional grids. For this method the principal directions of the local surface of L' are utilized as the 3D coordinates for the quantification.

A surface is fitted to the points on L' by using an approximation approach instead of interpolation. By doing this, the data becomes more robust to noise and outliers. The surface is then sampled on an uniform lattice (20 X 20). To avoid the effects that appear on the boundaries in the flanks of L' a larger region with r_2 is cropped, where $r_1 > r_2$. This bigger region is sampled on a bigger lattice and only the 20 X 20 central samples are concatenated to form a feature vector. After the application of the PCA, 200 feature vectors with dimension of 11 represent each face.

3.3.1.3 2D Feature Extraction

For 2D feature extraction, the SIFT algorithm is utilized. For each orientation of a keypoint a SIFT feature is extracted. After this, a 4 X 4 sample region is utilized to create orientation histograms, each one of them with 8 bins. SIFT features are robust against rotations

since, before the extraction, the gradients are rotated relative to the keypoints. In order to achieve robustness to illumination, the feature vectors are normalized to unit magnitude and large gradient magnitude suffer a threshold to a ceiling of 0.2, the vector is normalized again.

3.3.1.4 Matching

The same characteristics taken from the gallery faces are taken from the probe faces. After projecting those features into the PCA subspace, the next step is to calculate the similarity between them. This is done according to Equation 3.6.

$$e = \cos^{-1}(f_p^{\lambda}(f_g^{\lambda})^T) \tag{3.6}$$

being f_p^{λ} and f_g^{λ} the gallery and probe faces projected in the PCA subspace. If the two features are the same, then the value of e is zero. A matched feature is a feature in the probe that has the smallest error from the gallery. If there is more than one match, that one with the lowest value of e is considered. After all the features are matched a ordered list is created according to the values of e.

The keypoints from the matching features are projected in a xy-plane and meshed using the Delaunay triangulation. The next step is to project them back to the 3D space; with this, a 3D graph is generated. The list of matches is utilized in conjunction with the edges from the graph to create another graph; if the matches are right then both graphs are similar. The similarity between them are measured by:

$$\gamma = \frac{1}{n_\varepsilon} \sum_{i}^{n_\varepsilon} |\varepsilon_{pi} - \varepsilon_{gi}| \tag{3.7}$$

in which ε_{pi} and ε_{gi} correspond to the length of the edges from both graphs (probe and gallery face) and the value of n_ε is the total number of edges. The measure γ is invariant to face pose because the values of the lengths from the graphs edge's does not change even if the graph is rotated or translated. The Euclidean distance between two nodes after least square error minimization is another similarity measure, the outliers' nodes that have an error bigger than a threshold are throw away before calculating the distance. The threshold is the same of the distance for finding the keypoints.

For matching (MIAN; BENNAMOUN; OWENS, 2007) utilizes and define four measures of similarity, e, γ, the total number m of keypoint matches and the Euclidean distance d between two nodes. From those measures only m does not have a negative polarity. The probe face is matched to each face in the gallery resulting in four vectors S_q with the similarity measures of each comparison. Each vector is normalized with the min-max rule:

$$S'_q = \frac{S_q - \min(S_q)}{\max(S_q - \min(S_q)) - \min(S_q - \min(S_q))} \tag{3.8}$$

q is one of the similarity measures (e, γ, m and d). After the normalization the values of S'_m are subtracted from one 1 to reverse its polarity. The full similarity measure is calculated as:

$$S = K_e S'_e + K_m(1 - S'_m) + K_\gamma S'_\gamma + K_d S'_d \tag{3.9}$$

being K_q the confidence in each similarity measure, this can be achieved with:

$$K_q = \frac{\bar{S}_q - \min(S_q)}{\bar{S}_q - \min_2(S_q)} \tag{3.10}$$

having $\bar{S}_q$ the mean of S_q and $\min_2(S_q)$ the second minimum value of S_q.

3.3.2 Face Recognition under Varying Pose, Expression, Illumination and Disguise utilizing Kinect

The method proposed by Li et al. (2013) focused on face recognition utilizing scanners with low resolution. The proposed algorithm estimates canonical frontal view from non-frontal view and utilizes that information for achieving recognition. All of this is made using the nose tip as a reference point. This section gives an overview of this method.

3.3.2.1 3D Face Cropping and Pose Correction

The point cloud is translated in order to put the nose tip at the origin. Then, a sphere with radius of 8cm centered at the nose tip is utilized to crop the face. The result is a 6D point cloud (XYZ - RGB) containing only the surface of the face.

After cropping the face, the next step is to align the face to a reference model for pose correction. Since this is done utilizing the Iterative Closest Point (ICP) (BESL; MCKAY, 1992) algorithm, it would be unfeasible to align the probe face with all the faces on the gallery in search of the best alignment. To solve this problem the probe face is aligned with a reference model based on the Face Recognition Grand Challenge[1] (FRGC) (PHILLIPS et al., 2005) and UWA (MIAN, 2011) databases. This is necessary due to the high noise level of the Kinect data.

3.3.2.2 Symmetric Filling

After the pose correction, a mirrored image is generated by replacing the X points with their opposite numbers. If the Euclidean distance from a XY coordinate in the mirrored image to its closest neighbor in the original image is less than δ, then the point is removed. Otherwise, it will be added to the main point cloud.

The main objective of this step is to deal with missing data. The human face is not completely symmetric, but the difference in the two sides of the face is lower than the one caused by different identities. In their work Li et al. (2013) utilizes $\delta = 2mm$ (this value was found empirically).

[1] `http://www.nist.gov/itl/iad/ig/frgc.cfm`

3.3.2.3 Multi-Modal Sparse Coding

The symmetric filling process adds noise to the face model. Thus, to further improve the sample quality the model goes through a re-sampling stage. With this strategy the noise is attenuated and eventual holes are filled up. The method used in (LI et al., 2013) fits a smooth surface with the point cloud from the Kinect. This is done utilizing an approximation instead of interpolation.

For each face, 128 X 128 points are re-sampled in an uniform way. The RGB data is also re-sampled to the same XY positions using interpolation. This process generates four 128 X 128 RGB-D data. Those matrices are down-sampled to 32 X 32.

Based on (YIP; SINN, 2001), Li et al. (2013) utilize color information to improve recognition results. Normally the images are modeled in the RGB space, the problem with this is the high inter-component correlation that exists. With this in mind, the better solution is to transform the image to another space. In (LI et al., 2013) the images are transformed to the Discriminant Color Space (DCS).

The DCS space finds a set of linear combinations for each of the RGB channels. This is done to better separate the classes and decrease the intra-class variation (YANG; LIU; YANG, 2010). The DCS transform is applied to the texture from the image after the pre-processing steps.

For face recognition, a multi-modal Sparse Representation Classifier (SRC) is utilized. The SRC can correct small errors of missing data besides being robust to disguise. The classifier is applied to the depth data and to the DCS image separately. The main problem with this is that the DCS texture consists of three channels (one for R, other for G and other for B). For the sake of classification, those channels are stacked into one vector.

The scores are based on individual class reconstruction error for depth and texture. Those scores are normalized with the z-score technique and summed. The probe image is labeled with the image label that has the highest similarity score in the gallery. The sparse coding is formulated as a l_1 penalized linear regression problem, known as the LASSO problem (TIBSHIRANI, 1994; CHEN; DONOHO; SAUNDERS, 1998). For a detailed explication of the problem and its solutions the reader can refer to Tibshirani (2013).

3.3.3 Face Recognition Based on Local Low-Level Features

While the aforementioned methods (Li et al. (2013) and Kakadiaris et al. (2007)) focused on analyzing the deformation done on the face through expression or utilize methods that are robust to it, the method proposed by Lei, Bennamoun & El-Sallam (2013) focus on putting together low-level features taking only in consideration the nose (rigid), eyes and forehead (non-rigid). This section explains the core of this method.

3.3.3.1 Feature Extraction

This method utilized the FRGC v2 and BU-3DFE[2] (YIN et al., 2006) databases on experiments for validation and, because of holes and noise in the scans, the data need to go through a pre-processing. First, the values of all three coordinates of the vertices with discrepancy from their neighbors (x,y,z) are smoothed and, after that, a median filter is applied to all the surface. To fill in the holes, a bi-cubic filter interpolation is applied in all three coordinates.

Since the image is sampled at unordered locations, the next step is to convert them to range images imposing a fixed correspondence during the collection of features. To do so, the x and y are interpolated along the horizontal and vertical axis and z is determined as a pixel value. Those pixels are re-sampled at a distance of 1mm in both coordinates (x and y). Then, a Gaussian filter is applied to smooth the image.

The next problem to deal with is pose correction. The nose tip is positioned at the origin and is utilized to crop the outliers points, which are points located more than 80mm from the detected point. To do pose correction PCA is applied in the cropped points and their orientation is found. This is done repeatedly until there is no more pose change.

Utilizing empirical results Lei, Bennamoun & El-Sallam (2013) developed binary masks to crop the nose, mouth, eyes and forehead from the 3D scans. To reduce the size of the characteristics extracted, all cropped features are re-sampled at uniform intervals (2mm) and only the seeds are stored. Examples of the masks and the cropped features are given in the Figure 7.

After the pose correction, the range image is transformed into a point cloud. This is done by making the x and the y indices from a (x,y,z) matrix and z the depth data. Each region from the face is represented as multiple spatial triangles, which are made from a vertex selected utilizing the nose tip and two others random points chosen from the corresponding local surface region. From each triangle Lei, Bennamoun & El-Sallam (2013) proposes four types of geometric features, showed in Figure 8:

- A: The angle of two lines generated by two random points in the nose tip;
- C: The radius of the circle circumscribed in the triangle formed by the nose tip point and two random points;
- D: The distance of a line between two random points;
- N: The angle between a line defined between two random vertices and the z.

After the feature extraction they are normalized into (-1,+1) and grouped into histograms. The histogram generation is done by counting how many entries fall into each of the m bins. The Figure 9 shows the comparison of four histograms with a dimension of 180. It is important to note that even for different expressions the histograms from the same subject remain stable, while for another person it exhibits a visible difference.

[2] `http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html`

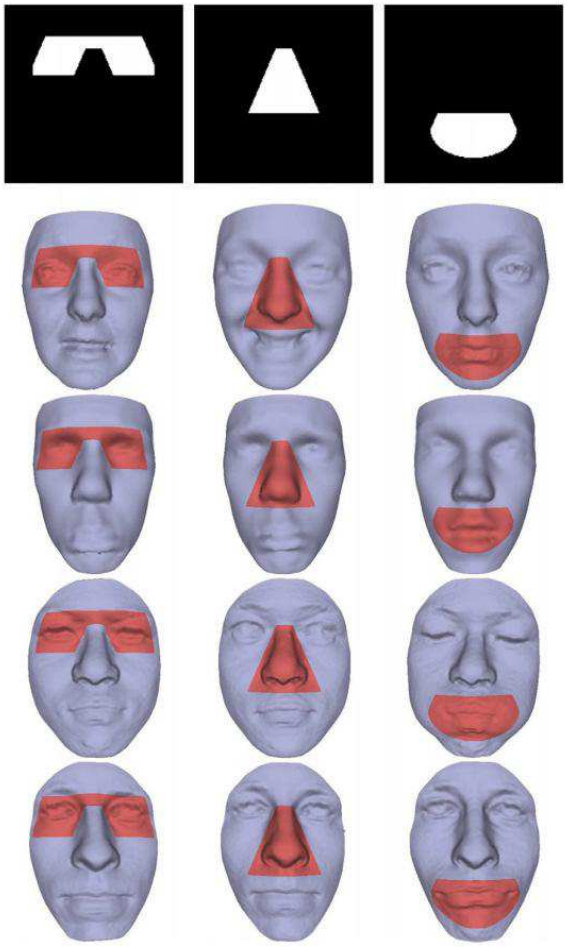


Figure 7 – Binary mask cropping the areas from the face (LEI; BENNAMOUN; EL-SALLAM, 2013).

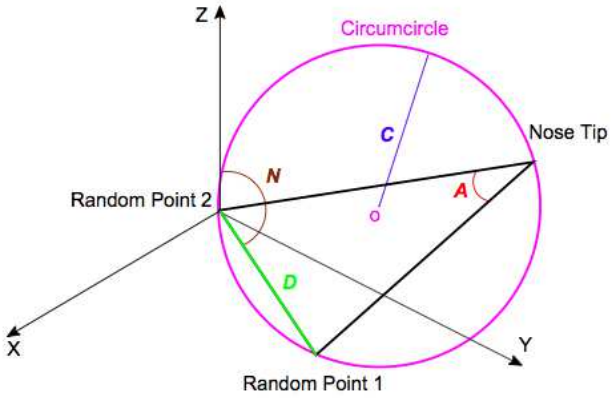


Figure 8 – An example of the four geometric features (LEI; BENNAMOUN; EL-SALLAM, 2013).

3.3.3.2 Support Vector Machines

On their work Lei, Bennamoun & El-Sallam (2013) utilizes SVM (Support Vector Machine) for classification. SVM maps the feature vector to a higher dimensional space and then finds an optimal hyper-plane that better separates two cluster of data. It does this by calculating the maximal margin.

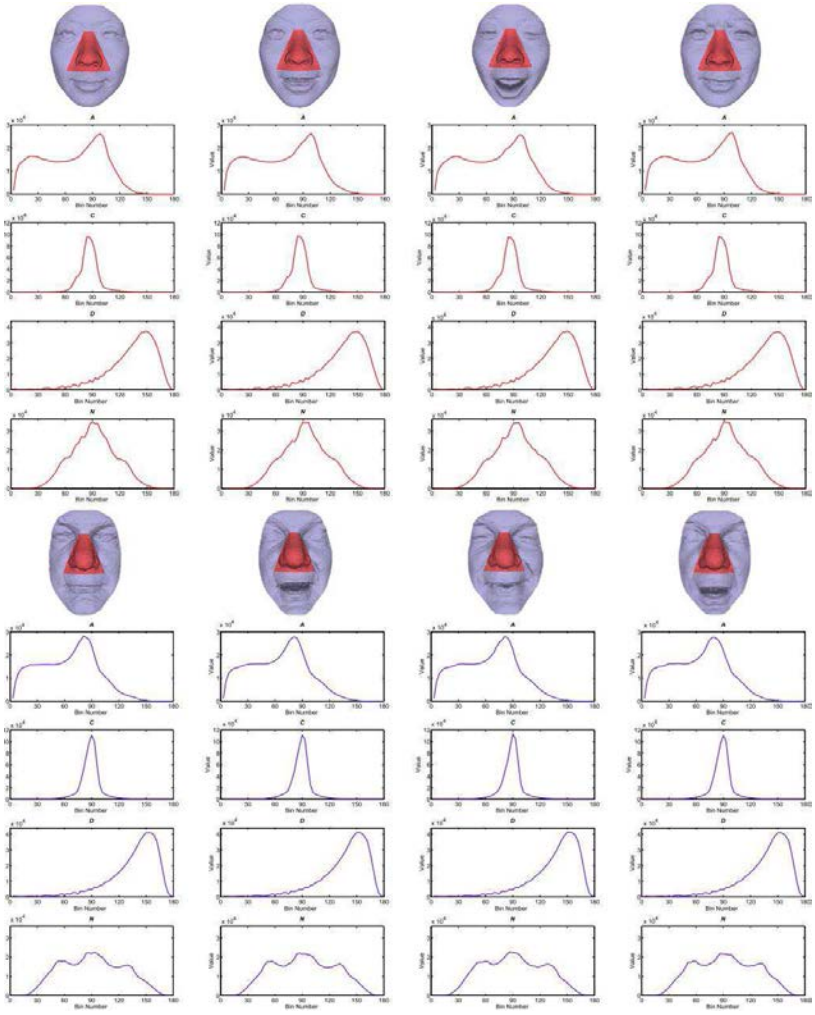


Figure 9 – Comparison of histograms generated from the same subject (LEI; BENNAMOUN; EL-SALLAM, 2013).

Given q labeled training samples $x_k \in \mathbb{R}^D, k = 1, \ldots, q$ belonging to two classes $y_k \in (+1, -1)$ SVM tries to find a hyper-plane solving the optimization problem:

$$\min_{\omega,b,\xi} \left(\frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l} \xi i \right) \tag{3.11}$$

$$s.t. y_i(\omega^T\phi(X_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \tag{3.12}$$

being $C > 0$ a penalty parameter of the error term, ω the coefficient vector, b a constant, and $\xi_i \geq 0$ a parameter for handling non-separable data. To facilitate the separation, the data is mapped into a function $\phi(x_i)$ that maps that data to a higher dimensional space. This function has a kernel form $K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$. For this method the non-linear Gaussian radial basis is chosen:

$$K(x_i, x_j) = exp(\gamma||x_i - x_j||), \gamma > 0 \tag{3.13}$$

in which both the γ and C need to be determined beforehand. The main objective is to define a hyper-plane separating data from two classes that maximizes the distance to the support vectors:

$$f(x) = \omega \cdot x + b, \tag{3.14}$$

$$\omega = \sum_{\forall x_i \in S} \alpha_i y_i x_i \tag{3.15}$$

being S a set of support vectors and α_i the trained weight of the support vectors. For classification, there are two schemes:

Distance-based: Computing the sign of $d(x)$ which is actually a function of the right side of Equation 3.14.

$$y(x) = \begin{cases} +1 & \text{sign(d(x))} = +1; \\ -1 & \text{sign(d(x))} = \text{-1}. \end{cases} \tag{3.16}$$

$$d(x) = \frac{\omega \cdot x + b}{||\omega||} \tag{3.17}$$

The sign of $d(x)$ is the classification of x and $|x|$ is the distance between x and the hyper-plane.

Probability-based: Probability based is carried out by mapping the real values of $f \in [+\infty, -\infty]$ to a probability distribution. This is done by training a sigmoid function:

$$P(y = +1|f(x)) = \frac{1}{1 + \exp(Af + B)} \tag{3.18}$$

being A and B two parameters estimated utilizing the training set, P the mapped probability, and y the label. If P is high there is a higher probability to that data belonging to class $+1$ and vice-versa.

Since SVM was created only for binary classification, to solve multi class problems it is necessary to utilize an approach as proposed by Lei, Bennamoun & El-Sallam (2013), which utilizes the one-vs-all method. This approach trains SVMs to classify k classes and each of the SVMs are responsible to separate the sample from one class (+1) to all the other samples (classified as -1).

3.3.3.3 Feature-Level Fusion

Since there is two kinds of features being analyzed (rigid and semi-rigid regions), it is necessary to fuse both biometric characteristics. Feature-level requires that the features are fused directly. This is done by concatenating the histograms for each region. For classification, q subjects are trained and the values of C and γ are defined utilizing a grid-search algorithm with 5-fold cross-validation.

Let x be the probe scan, then the distance $d_k(x)$ is the kth element of the similarity vector $\vec{d}(x)$. The vectors are normalized with the min-max rule:

$$\vec{d'}(x) = \frac{\vec{d}(x) - \min(\vec{d}(x))}{\max(\vec{d}(x) - \min(\vec{d}(x))) - \min(\vec{d}(x) - \min(\vec{(d)}(x)))} \tag{3.19}$$

being $\vec{d'}(x)$ the normalized distance vector. In verification mode, given a classification threshold η

$$\text{Accept if } d'_k(x) + \eta > 0$$

$$\text{Reject if } d'_k(x) + \eta \leq 0$$

For identification mode the label $y(x)$ is computed in the following form:

$$y(x) = \arg \max_{1 \leq k \leq q} (d'_k(x)) \tag{3.20}$$

3.3.3.4 Score-Level Fusion

With score-level fusion, it is necessary to train two sets of SVM (S-SVMs for the semi-rigid features and R-SVMs for the rigid features). The score-level fusion is the weighted sum of the individual outputs from each SVM set.

The training for each set is the same and generates q SVMs that recognize q different subjects. Because of this is necessary to estimate a pair of sigmoid function parameters $(A_k, B_k), k = 1, \ldots, q$ as follows:

- Use the training set to train q SVMs and find the optimal C and γ as already described;
- Give all the labeled samples as a input to the q SVMs to generate a set of $(f_i, y_i)_k$, i is the ith sample, f_i is computed by the equation 3.14 as the decision value and y_i is the estimated labels between -1 and +1;
- The algorithm (PLATT, 1999) is applied in $(f_i, y_i)_k$ to estimate (A_k, B_k);
- Repeat the two steps above to obtain q pairs of (A, B).

For a probe scan x two probability vectors $\vec{p}^s$ and $\vec{p}^r$ are created (which correspond to S-SVMs and R-SVMs respectively) and mapped by their sigmoid functions. Each of them have q elements and the sum of them results in 1. A higher value at the position k means a higher chance for the probe to belong to the kth class.

Even though both SVMs are trained in the same way, the results with the S-SVMs are more reliable. This indicates the need for different weights for the fusion of those scores. Having

$\vec{w}^s$ and $\vec{w}^r$ as the weight vectors for S-SVMs and R-SVMs, respectively, it is possible to calculate them as:

$$\vec{w}_k^s = \sum_{i=1}^{n}\sum_{k=1}^{q}\frac{p_k^s(x_i)}{q}\cdot\sum_{i=1}^{n}p_k^s(x_i) \tag{3.21}$$

$$\vec{w}_k^r = \sum_{i=1}^{n}\sum_{k=1}^{q}\frac{p_k^r(x_i)}{q}\cdot\sum_{i=1}^{n}p_k^r(x_i) \tag{3.22}$$

and for the final weight vector $\vec{w} = w_1, \ldots, w_k, \ldots, w_q$:

$$w_k = \frac{w_k^s}{(w_k^s + w_k^s)} \tag{3.23}$$

And the fusion of the probabilities:

$$\vec{p}(x) = \vec{w} * \vec{p}^s(x) + (1 - \vec{w}) * \vec{p}^r(x) \tag{3.24}$$

In verification mode, given a threshold η:

$$\text{Accept if} \, p'_k(x) > \eta$$

$$\text{Reject if} \, p'_k(x) \leq \eta$$

in which p'_k is the probability vector of the probe facial scan x belonging to the subject k.

In identification mode, the class $y(x)$ is computed as:

$$y(x) = \arg\max_{1\leq k\leq q}(p'_k(x)) \tag{3.25}$$

The label of the class is the largest $p'_k(x)$

3.3.4 RGB-D Face Recognition utilizing Kinect

In their work Goswami et al. (2013) utilize Kinect data for face recognition. They focus on utilizing depth and RGB data for recognizing different individuals, this is done extracting entropy and salience maps from the data. Histogram of Oriented Gradients (HOG) are utilized as the face descriptor. In this section this method is described.

3.3.4.1 Extraction of Entropy and Visual Salience Maps

The entropy maps are applied in both type of data (depth and RGB). The depth maps generated by the Kinect does not have big variations, because of this Goswami et al. (2013) affirms that the information seems insignificant for feature extraction. Entropy is utilized to increase such variation, and then increase the results of the feature extraction.

In order to measure the entropy H of a random variable x the following formula is utilized:

$$H(x) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i) \tag{3.26}$$

being $p(x_i)$ the value of the probability density function for x_i.

For extracting the entropy maps, the image is represented by a pair of intensity functions $[I_{rgb}(x, y), I_d(x, y)]$ with size $M \times N$, both of them defined over the same (x, y), with $x \in [1, M]$ and $y \in [1, N]$. Four patches are extracted from the images: P_1 and P_2 from the RBG and P_3 and P_4 from the depth data. For P_1 and P_3 the patches have size of $\frac{M}{2} \times \frac{N}{2}$ centered at $[\frac{M}{2}, \frac{N}{2}]$, while for P_2 and P_4 the size is $\frac{3M}{4} \times \frac{3N}{4}$ centered at $[\frac{M}{2}, \frac{N}{2}]$.

Utilizing P_1, P_2, P_3, P_4 four entropy maps are computed. This is done with:

$$E_i = H(P_i), i \in [1, 4] \tag{3.27}$$

Besides the entropy, the saliency map is also extracted from the RGB image. The saliency is not extracted from the depth maps because the methods are made specifically for visual images and, since the depth is a measure of distance and not intensity information, the direct application can result in irregular output.

Visual salience is the capability of an image area to attract attention (DESIMONE; DUNCAN, 1995). It is defined as an intensity function $S(\dot{})$ over an image $I(x, y)$. This function has the role of mapping individual pixels to a proportional value of saliency. In their work (GOSWAMI et al., 2013) implement the saliency map extraction based on a MATLAB code[3] and (ITTI; KOCH; NIEBUR, 1998).

3.3.4.2 Histrogram of Oriented Gradient

The Histograms of Oriented Gradient (HOG) descriptor is based on the magnitude and orientation of the gradients in an image. In their work (GOSWAMI et al., 2013) apply the descriptor to both Entropy and Salience maps. For each face there are five different histograms, two from the patches E_1, E_2 (entropy maps based on RGB images) and two from E_3, E_4 (entropy maps based on depth maps). The last histogram is based on the visual saliency map (E_5).

The final descriptor is the ordered concatenations of the five HOG histograms, as follows:

$$F = [F_1, F_2, F_3, F_4, F_5] \tag{3.28}$$

being F_n the HOG from the E_n map.

For the classification the authors choose to utilize the Random Decision Forest (RDF). They utilize grid search on a group of training samples in order to discover the optimal parameters for the classification problem, but the values are not given in the original work.

[3] `http://www.vision.caltech.edu/~harel/share/gbvs.php`

3.3.5 Comparison of the 3D Face Recognition Methods

In their work, Lei, Bennamoun & El-Sallam (2013) compared the results obtained by their method with the ones obtained by Mian, Bennamoun & Owens (2007), since both utilize the FRGC v2 database in their experiments. The FRGC v2 contains 4.007 3D scans of 466 different persons, various facial expressions, with 57% males and 43% females. The age distribution is 65% of 18-22 years old subjects, 18% of 23-27 and 17% of 28 years old or more. The UH dataset is an extended FRGC v2 which contains 884 3D faces (KAKADIARIS et al., 2007). The base is more challenging because the subjects use facial accessories and make extreme facial expressions. The results of tests utilizing 466 individuals are shown in Table 4.

Table 4 – Comparison between methods of 3D face recognition (LEI; BENNAMOUN; EL-SALLAM, 2013).

Method	Modalities	No. of Probes	Rates(%)
(MIAN; BENNAMOUN; OWENS, 2007)	2D + 3D	1597(466)	96.6
(LEI; BENNAMOUN; EL-SALLAM, 2013)	3D	1000(466)	97.6

Lei, Bennamoun & El-Sallam (2013) also tested their method with the BU-3DFE database. The BU-3DFE is a database with 100 individuals. Each of those individuals has 25 ear-to-ear scan and only one in neutral expression, the other scans are divided into six expressions: Happiness, Anger, Fear, Disgust, Sadness and Surprise. The expressions are divided into 4 levels of intensity: 1 and 2 were considered low level, while 3 and 4 were considered of high intensity level.

The Figure 10 shows the CMC curves for face identification in the FRGC v2 database (left column) and the BU-3DFE (right column). It is possible to observe that the feature-level fusion has better performance in all the results in this figure, even when comparing different levels of expression intensity.

The Figure 11 shows the ROC curve for face verification in the FRGC v2 database (left column) and the BU-3DFE (right column). Both experiments utilize feature-level fusion and score-level fusion. In the ROC curve it is possible to observe that, even though the feature-level fusion has better performance, both fusion types are not very far from the other.

The method proposed by Li et al. (2013) was not assessed on FRGC v2 because it was proposed to utilize Kinect data as input, so it is not feasible to compare it with the two other methods. This method was validated with the CurtinFace database[4] that contains over 5000 images of 52 subjects with variations in pose, illumination, facial expression and sunglasses. All of this is captured with a Kinect sensor.

The Table 5 shows the results of recognition with poses and expression variation, whilst the Table 6 shows the results of recognitions with illumination and expression variation. The results without the use of symmetric filling makes it obvious the need of it. If there is a pose of $\pm 60^{o} yaw$ some information of the face is totally lost, but with the symmetric filling, part of this data can be reconstructed to some extent. The symmetric filling is effective even when there is no pose variation. This can be seen in the results of illumination and expression variation.

In Table 5 it is possible to see that the fusion between the depth map and the texture

[4] `http://impca.curtin.edu.au/downloads/datasets.cfm`

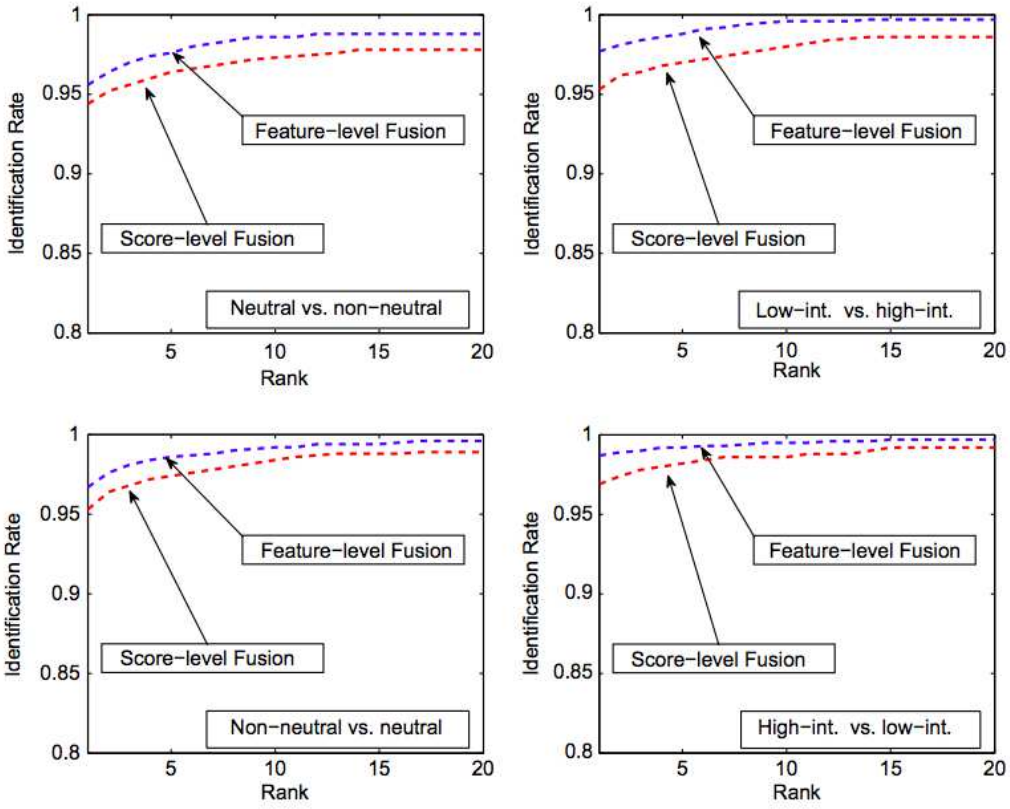


Figure 10 – CMC curves comparing the identification results (LEI; BENNAMOUN; EL-SALLAM, 2013).

Table 5 – Results for recognition tests in the database CurtinFaces with the method proposed by (LI et al., 2013). In this test there are pose X facial expression variations.

Pose	Depth Map without Symmetric Filling	Texture Map without Symmetric Filling	Fusion	Depth Map with Symmetric Filling	Texture Map with Symmetric Filling	Fusion
Frontal	100	100	100	100	100	100
$\pm 30°yaw$	49.5	98.1	93.6	88.3	99.8	99.4
$\pm 60°yaw$	14.9	80.4	55.1	87.0	97.4	98.2
$\pm 90°yaw$	1.0	39.4	14.4	74.0	87.3	84.6
$\pm 60°pitch$	77.2	91.3	90.9	81.6	89.1	92.8
Average	46.2	87.6	77.0	85.4	95.0	96.3

Table 6 – Results for recognition tests in the database CurtinFaces with the method proposed by (LI et al., 2013). In this test there are illumination X facial expression variations.

Illumination	Depth Map without Symmetric Filling	Texture Map without Symmetric Filling	Fusion	Depth Map with Symmetric Filling	Texture Map with Symmetric Filling	Fusion
Front	89.1	96.8	98.4	92.5	97.1	98.9
Back	89.4	96.6	97.6	93.8	96.5	98.6
Low Ambient	87.2	91.0	95.8	91.3	91.0	97.1
Average	88.8	95.6	97.6	92.8	95.6	98.4

map, when not utilizing symmetric filling, can decrease the performance. In all cases the depth maps presented worst performance than the texture map. When utilizing symmetric filling in

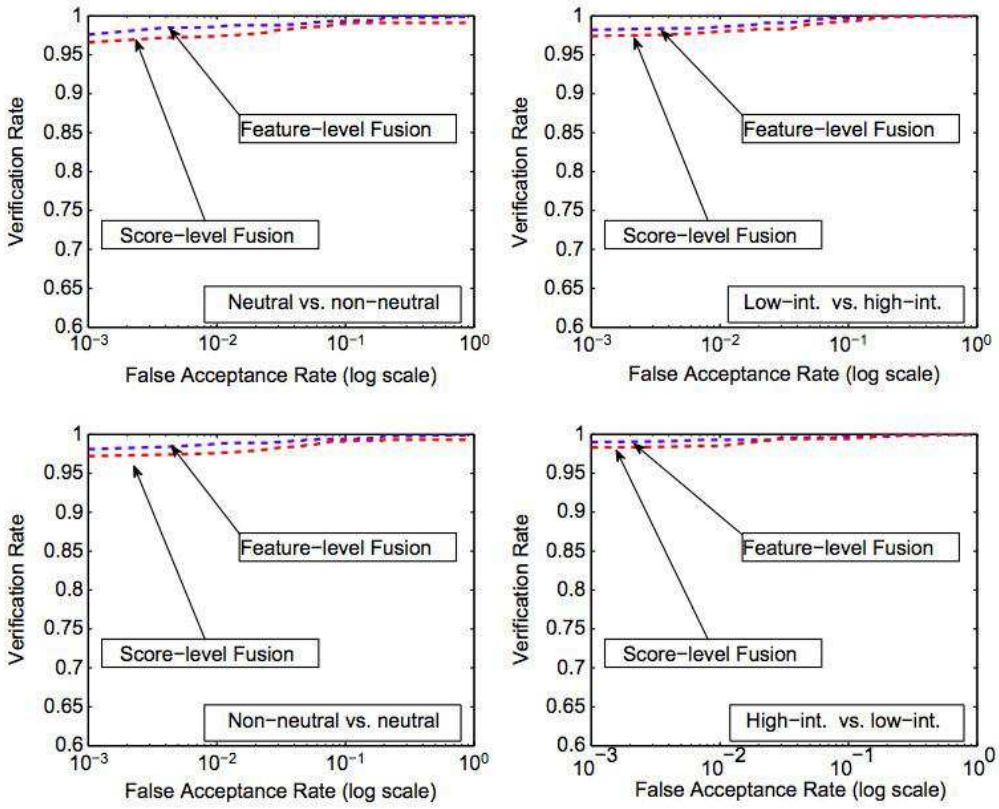


Figure 11 – ROC curves comparing the verification results (LEI; BENNAMOUN; EL-SALLAM, 2013).

the presence of rotation the fusion of both data types seems to increase in most of the cases, only with 30 degrees of yaw rotation that there is a loss of performance.

In Table 6 the fusion seems to benefit the performance in all cases. Once more the symmetric filling proves to be valuable, since it increased the performance in all the cases of the second experiment. It is important to note that the Depth Maps is the data source that seems to gain more benefit from the symmetric filling.

3.4 Local Binary Pattern

The Local Binary Pattern (LBP) was originally introduced by Ojala, Pietikäinen & Harwood (1996) and it consists in analyzing the difference between a pixel and its 3 X 3 neighborhood. Given a central pixel (x_c, y_c) the LBP operator can be given as (RODRIGUEZ; MARCEL, 2006):

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \tag{3.29}$$

with i_c as the grey value at the center pixel, i_n the value of the neighborhood pixel and $s(x)$ defined as:

$$s(x) = \begin{cases} 1 & \text{if x} \geq 0; \\ 0 & \text{if x} < 0. \end{cases} \tag{3.30}$$

The Figure 12 shows a practical example of how the code is generated.

Later Ojala, Pietikainen & Maenpaa (2002) proposed an extension to the original LBP method. Instead of using only a 3 X 3 neighborhood, with this new approach it has become possible to indicate (P, R), being P the quantity of sampling points in a circle of radius R. The Figure 13 illustrates the $LBP_{(4,1)}$ circular neigborhood of a pixel.

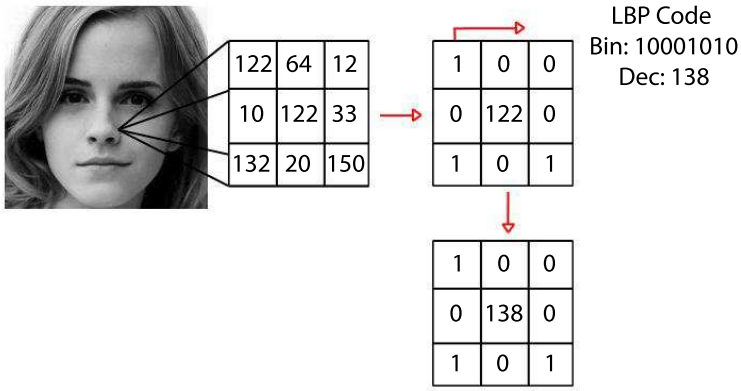


Figure 12 – Example of how to calculate the LBP operator.

If the coordinates of the central pixel are $(0, 0)$ the coordinates of a neighbor g_p are given by:

$$\left(-R\sin\left(\frac{2\pi p}{P}\right), R\cos\left(\frac{2\pi p}{P}\right)\right) \tag{3.31}$$

The values that does not fall in the center of a pixel are estimated by bilinear interpolation.

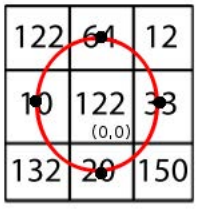


Figure 13 – Example of a $LBP_{(4,1)}$ neigborhood of a pixel. The black dots are the sampling points in the red circle.

3.4.1 3D Local Binary Pattern

The LBP operator only takes in consideration the signal of the comparison between a region and its kernel. The operator was originally proposed for texture description and cannot deal with the behavior of depth values. One can observe that if two central points on different

samples have highest (or lowest) depth values than their neighbors, they will have the same operator value, even if they are from different subjects (HUANG; WANG; TAN, 2006). This would be common on points belonging to the nose tip of a face subject, for instance.

To deal with situations like this Huang, Wang & Tan (2006) proposed the 3D Local Binary Patterns (3DLBP). This variation of the original operator considers not only the signal of the difference, but also the absolute depth difference. Huang, Wang & Tan (2006) state that more than 93% of all depth differences (DD) with $R = 2$ are smaller than 7. Due to this property the absolute value of the DD is stored in three binary units ($i_2 i_3 i_4$). Therefore, it is possible to affirm:

$$|DD| = i_2 \cdot 2^2 + i_3 \cdot 2^1 + i_4 \cdot 2^0 \tag{3.32}$$

There is also i_1, a binary unit defined by:

$$i_1 = \begin{cases} 1 & \text{if DD} \geq 0; \\ 0 & \text{if DD} < 0. \end{cases} \tag{3.33}$$

Those four binary units are divided into four layers and, for each of those layers, four decimal numbers are obtained: P_1, P_2, P_3, P_4. The value of the P_1 has the same value as the original LBP. For matching, the histograms of the local regions (P_1, P_2, P_3, P_4) are concatenated. The Figure 14 shows the process for the generation of the 3DLBP, given an image.

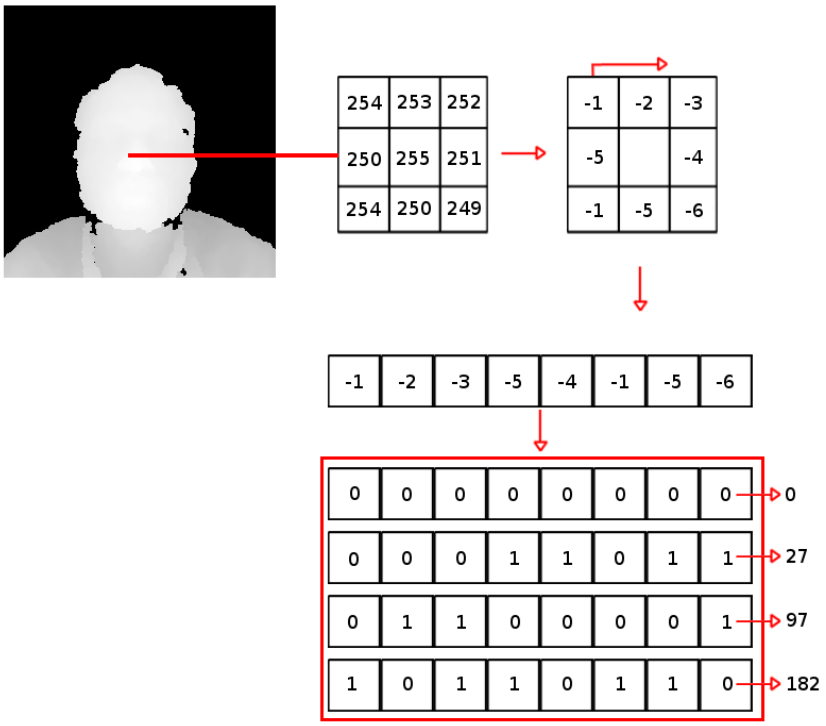


Figure 14 – The full process of the 3DLBP proposed by (HUANG; WANG; TAN, 2006). Each of the differences is encoded into the layers (layer 2, 3 and 4) and the signal into the layer 1.

3.4.2 Gradient-LBP

In their work Huynh, Min & Dugelay (2012) proposed a new variation from the LBP operator called Gradient-LBP. In this method, each orientation of the depth difference is taken in consideration, with this there is a depth difference image for each orientation. An $LBP_{(8,1)}$ generates eight orientations and, for each one of them, there is a depth difference image.

The oriented depth difference at a position (x_c, y_x) is defined by:

$$ODD^{P,R,p}_{p=0...P-1} = max(min(g_p - g_c, 7), -8) \tag{3.34}$$

Being ODD the Oriented Depth Difference at pixel (x_c, y_x) in the orientation p, g_p the depth value at point p, and g_c the depth value at the central point. The value of the ODD will stay between -8 and 7. The original work is focused in gender recognition, but it would be interesting to study the application of the aforementioned method also for face recognition.

3.5 Histogram of Averaged Oriented Gradients

Galoogahi & Sim (2012) proposed a method for inter-modality Face Sketch Recognition. The main goal of the method is to reduce the gap made by the different modality between face photos and sketches. The method is based on a new gradient orientation descriptor, Histogram of Averaged Oriented Gradients (HAOG).

This gap is generated by the difference of visual information that can be seem in a photo and a sketch. A face has several components (e.g. eyes, eyebrows, lips) that has a strong relations to each other in a spatial configuration (GALOOGAHI; SIM, 2012). The general shape of the face and its spatial configuration are the meaningful visual information and not the size of facial components. With this in mind, it is possible to affirm that the amount of shape information from a photo and a sketch is the same; the face shape is not involved in the modality gap.

While the face shape does not generates modality gap, the same cannot be said for the texture. Sine texture is related to face appearance it is deeply related with the how the modality gap presents itself. Face appearance has coarse and fine textures, belonging to facial components and facial skin.

Boundaries of the facial components with high contrast are coarse textures, while the low contrasts from the face skin (e.g. flaws, moles) are fine textures. It is possible to explore the problem separately for each type of textures. Coarse textures are vital for artists to draw sketches but fine textures details can be lost in the sketch.

Since the discriminative power of the extracted orientation gradients are related with the modality gap and only the fine textures are responsible for the gap it can be concluded that the more robust way to deal with inter-modality recognition is to use only coarse textures for feature extraction.

Since it is not feasible to separate both the texture types, the best way to deal with this situation is to use both of them but emphasizing on the coarse textures. One way to achieve this is to vote square magnitudes of gradient into histogram of orientations.

The HAOG descriptor is computed in three main steps. First the grayscale image has a $S \times S$ window sliding over it, starting in the left upper corner and ending at the right lower corner. For each pixel in the local patches the $\bar{\rho}$ and $\bar{\varphi}$ are defined. For each patch a histogram with b bins is defined, the histogram is quantified by accumulating $\bar{\rho}$ at the bin which $\bar{\varphi}$ fell into. The final descriptor is the concatenation of all the histograms for the patches.

For $\bar{\rho}$ and $\bar{\varphi}$ first it is needed the gradient vector for the image. Given an image in grayscale $I(x, y)$, the gradient vector can be defined as:

$$\begin{bmatrix} g_x(x,y) \\ g_y(x,y) \end{bmatrix} = \begin{bmatrix} \frac{\partial I(x,y)}{\partial x} \\ \frac{\partial I(x,y)}{\partial y} \end{bmatrix} \tag{3.35}$$

The values of ρ and φ can be calculated with:

$$\rho = (g_x^2 + g_y^2)^{0.5} \tag{3.36}$$

$$\varphi = \tan^{-1}\left(\frac{g_y}{g_x}\right) \tag{3.37}$$

The main problem is that it is unfeasible to calculate directly the averaged oriented gradient, opposite gradients at both sides of an edge can cancel each other. To deal with this problem (KASS; WITKIN, 1987) proposed to double the gradient angles before averaging. For defining $\bar{\rho}$ and $\bar{\varphi}$ first is needed to define the average squared gradient for each pixel in a local neighborhood in a window W, this is done by:

$$\begin{bmatrix} g_{sx} \\ g_{sy} \end{bmatrix} = \begin{bmatrix} \rho^2 \cos 2\varphi \\ \rho^2 \sin 2\varphi \end{bmatrix} = \begin{bmatrix} \rho^2(\cos^2\varphi - \sin^2\varphi) \\ \rho^2(2\sin\varphi\cos\varphi) \end{bmatrix} = \begin{bmatrix} g_x^2 - g_y^2 \\ 2g_y g_x \end{bmatrix} \tag{3.38}$$

$$\begin{bmatrix} \bar{g}_{sx} \\ \bar{g}_{sy} \end{bmatrix} = \begin{bmatrix} \sum_W g_{sx} \\ \sum_W g_{sy} \end{bmatrix} = \begin{bmatrix} \sum_W (g_x^2 - g_y^2) \\ \sum_W 2g_y g_x \end{bmatrix} \tag{3.39}$$

Then, $\bar{\rho}$ and $\bar{\varphi}$ are defined by:

$$\bar{\varphi} = \tan^{-1}\left(\frac{\bar{g}_{sy}}{\bar{g}_{sy}}\right), \bar{\varphi} \in [-\pi, \pi) \tag{3.40}$$

$$\bar{\rho} = \sum_W (g_{sy}^2 + g_{sx}^2)^{0.5} = \sum_W \rho^2 \tag{3.41}$$

3.6 Conclusion

As we could observe in this chapter, even with noise and low resolution data, as those outputted by Kinect, it is possible to propose methods that can recognize different individual faces.

The method proposed by Li et al. (2013), for instance, shows an efficient way to deal with those problems and reaches high recognition results. Unfortunately, the probability density function for the impostor scores and the EER are not shown in their work.

The method proposed by Goswami et al. (2013), which utilizes the entropy map for each channel (RGB-D) for recognition, reaches 80% of correct recognition at Rank 1.

The other two methods described in this chapter also presented high recognition rates in the experiments conducted by their authors. However, they were proposed for high resolution scanners data. Then, it is not possible to affirm that such methods would provide the same performance utilizing Kinect data. When dealing with depth data the methods, in general, need adaptation, as can be seen with the LBP operator.

One approach to get good results dealing with noisy and low-resolution data, likewise Kinect data, would be to combine different methods, so one can compensate the fragilities of others, and vice-versa. In our work, we propose a method for 3D face recognition based on the fusion of different face descriptors, in order to improve the recognition rates.

Chapter 4

Proposed Method

In this chapter the proposed method for 3D facial recognition utilizing Kinect is presented. This method does a weighted score fusion from the 3DLBP (3D Local Binary Pattern) and HAOG (Histogram of Averaged Oriented Gradient) features, as illustrated the Figure 15, in order to provide a better performance. The next sections of this chapter are dedicated to describe each stage of the proposed method.

4.1 Face Normalization

The nose tip is the principal landmark utilized in our method for the face normalization. All faces must have their nose tip translated to the origin. This is necessary, especially for the symmetric filling step.

The depth maps are cropped in a rectangular region centered at the nose tip. The width and the height of the cropped area is the distance between both eyes. The face features are extracted from the cropped area.

4.2 Feature extraction

As shown in the Chapter 3 there are some different methods to deal with 3D data in order to extract features from the face. The new method proposed in this dissertation aims to fuse information extracted from the depth maps, utilizing 3DLBP and HAOG features, as described in Sections 3.4.1 and 3.5, respectively.

One can observe that these two face descriptors are represented as histograms.

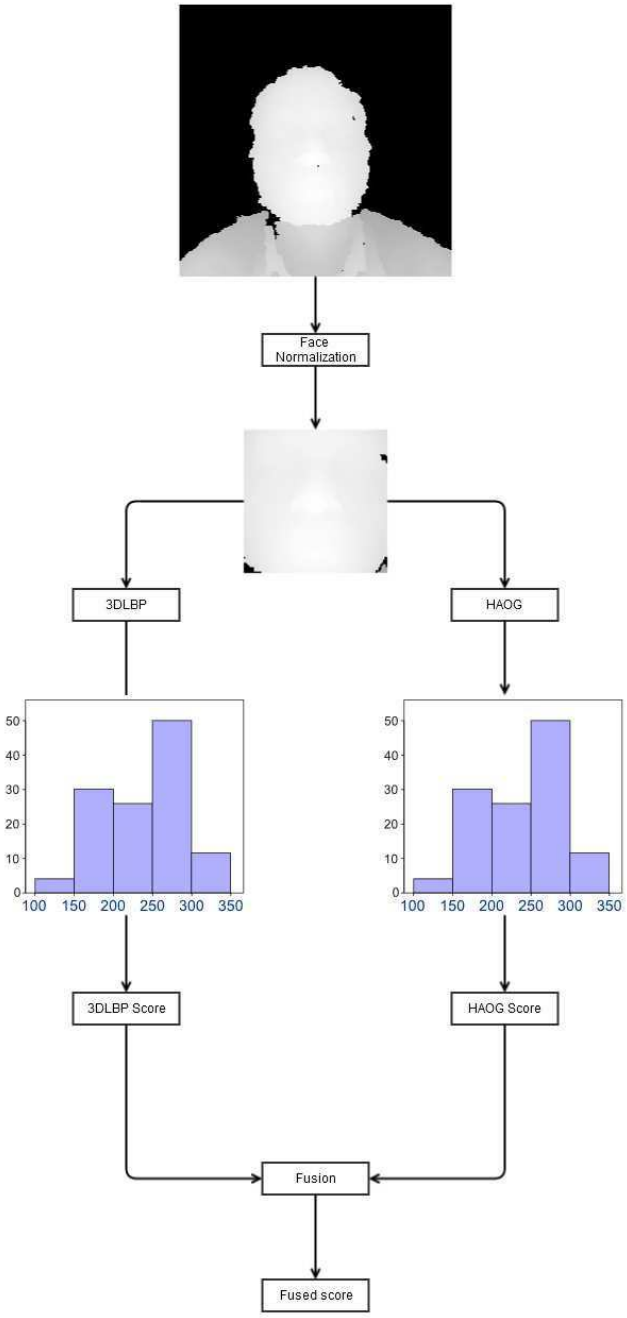


Figure 15 – Diagram of the proposed method for 3D face recognition using Kinect data. The proposed method combines the 3DLBP and HAOG methods in order to provide a better performance.

4.3 Fusion Strategy

In the proposed method for 3D face recognition using Kinect, the fusion in carried out in the score-level (JAIN; ROSS; PRABHAKAR, 2004). Thus, the final classification score is given by:

$$FS = 3DLBPSC * w_1 + HAOGSC * w_2 \tag{4.1}$$

being FS the final score, $3DLBPSC$ the score from the 3DLBP method, w_1 the weight for the 3DLBP score, $HAOGSC$ the score from HAOG classification, and w_2 the weight for the HAOG

classification. For a more detailed description on score-level fusion the Section 3.3.3.4 can be consulted. Both the $3DLBPSC$ and $HAOGSC$ are normalized before the fusion.

4.4 Generation of Depth Maps from the Cloud Points

In our method, aiming to increase the robustness of the 3D face recognition from kinect data, we also generate depth maps from the cloud points, outputted by the Kinect device.

In order to generate a depth map from the cloud points, a circular region with radius R is cropped centered at the nose tip. Then, the cropped image goes through the symmetric filling process, as described in Section 3.3.2.2. Finally, the resulting face image is fitted to a smooth surface using an approximation approach, which is done with an open source code[1] written in MATLAB. The result of this process is a 100 X 100 matrix, stored in an image file. Since the image file cannot contain decimal values, all the depth values are rounded.

The Figure 16 shows an example of the two types of depth maps utilized in the proposed method: (a) the one obtained directly from the Kinect, and (b) the one generated with the cloud points after the symmetric filling process.

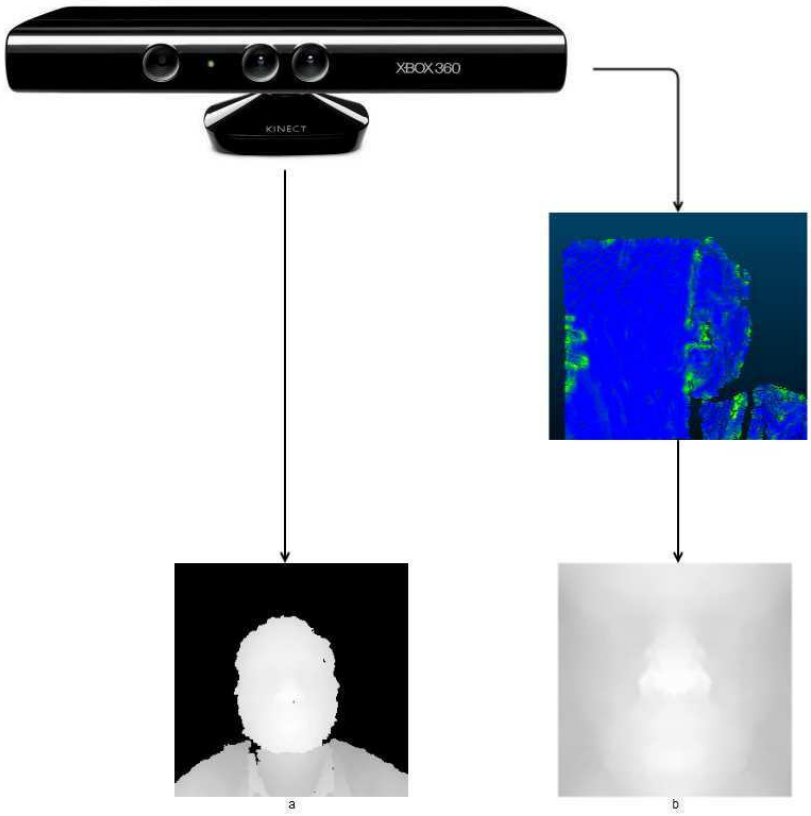


Figure 16 – Two types of depth maps utilized in the proposed work. a) Depth map generated directly by the Kinect device. b) Depth map generated from the cloud points outputted by the Kinect device after the symmetric filling and approximation processes.

[1] http://mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit

This pre-processing step is only utilize when there is obstructed faces in the samples being analyzed. It is illustrated in the Figure 17.

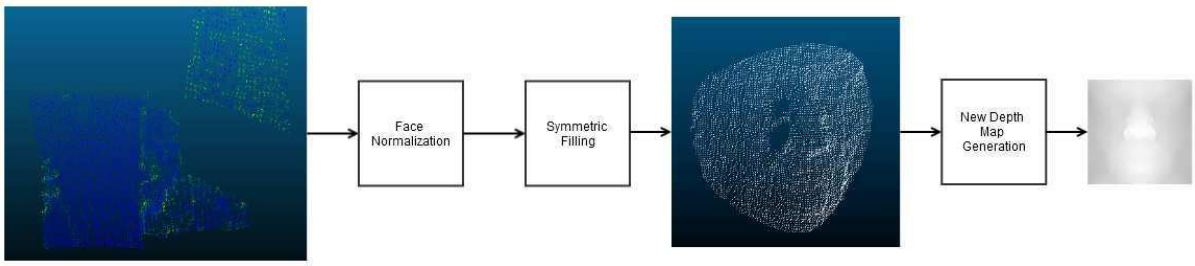


Figure 17 – The generation of new depth maps based on the cloud points

Chapter 5

Experiments and Results

In order to assess the performance of the proposed method, experiments on a public Kinect face database (EURECOM Kinect Dataset (HUYNH; MIN; DUGELAY, 2012)) were conducted. Details of the face database and the results obtained with these experiments are presented in this chapter.

All the experiments were made utilizing a personal computer with the following configuration:

- Processor: Intel i7-3530M (2.90 GHz);
- RAM: 8 GB;
- HD: 1 TB;
- Operational System: Microsoft Windows 8 64 bits.

5.1 EURECOM Kinect Dataset

The EURECOM Kinect Dataset (HUYNH; MIN; DUGELAY, 2012) is composed of data obtained from 52 subjects, 14 females and 38 males, using a Kinect device.

In this dataset, there are two sets of images captured in the interval of 15 days, and for each subject there are nine images, with the following characteristics: neutral face, smiling, open mouth, lightning, occlusion of the eyes, occlusion of the mouth, occlusion of the right side of the face, left profile, and right profile.

For each face, there are three different data sources: depth map (in bitmap depth images and text files with all the values sensed by the Kinect), RGB image, and 3D .obj files. Each format has the annotation for: the left and right eyes, the tip of the nose, left and right side of the mouth and the chin.

The Figure 18 shows a set of images from the database and the Figure 19 shows the annotated landmarks.

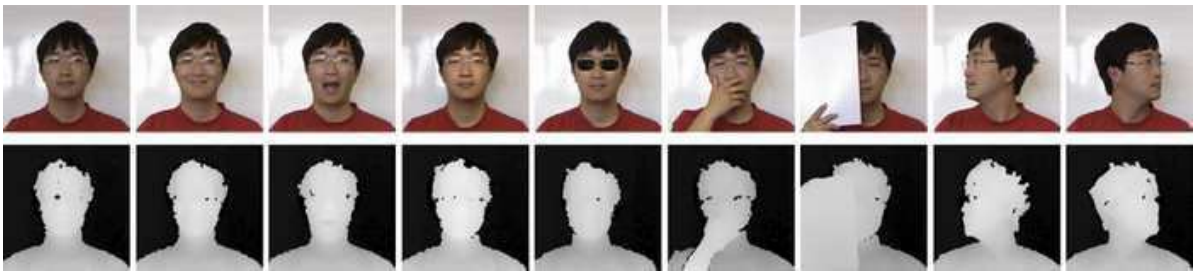


Figure 18 – A set of images from a subject in the EURECOM database, in which it is possible to see different poses and facial expressions in the RGB (top row) and the depth map (bottom row) images.

Figure 19 – An example of annotated face landmarks in the EURECOM database: left and right eyes, the tip of the nose, left and right side of the mouth and the chin.

The sessions took place in the lab of the EURECOM Institute. The subjects were filmed twice for each session and, during the footage, they were asked to do different facial expressions. Each person stood at approximately one meter far from the Kinect device and each image was cropped to a 256X256 image centered at the face.

5.2 Experiments Protocol and Performance Measurement

We have carried out three experiments in this work in order to assess the propose method for 3D face recognition using Kinect. In the first one, it was utilized a reduced set of images from the EURECOM Kinect Dataset, which was called Set 1. In the second experiment, it was utilized a set of images, called Set 2, composed by all images from the Set 1, plus the images of the EURECOM Kinect Dataset in which the subjects have the right side of their faces occluded. One can see the Section 5.3 for more details about the Sets 1 and 2.

In order to calculate the probability density functions (pdf) for the genuine and impostor scores, we carried out:

- Genuine comparisons: Each face sample in the probe set was compared against all the other face samples in the gallery set of the same person;
- Impostor comparison: Each face sample in the probe set was compared against all the other face samples in the gallery set of all other persons.

From the probability density functions for the genuine and impostor scores, we calculated the False Acceptance Rate (FAR) and the Genuine Acceptance Rate (GAR), according to the following equations:

$$FAR = p(s \geq \eta|\omega_0) = \int_{\eta}^{\infty} p(s|\omega_0)ds \tag{5.1}$$

$$GAR = p(s \geq \eta|\omega_1) = 1 - FRR(\eta) \tag{5.2}$$

$$FRR = p(s \geq \eta|\omega_1) = \int_{-\infty}^{\eta} p(s|\omega_1)ds \tag{5.3}$$

being η the system threshold, ω_0 the impostor class, ω_1 the genuine class, p the probability function, and s the matching score.

Since a biometric system can operate at different threshold values, the FAR and GAR at different values of threshold are measured and summarized in the form of a Receiver Operating Characteristics (ROC) curve (JAIN; ROSS; NANDAKUMAR, 2011).

The analysis of a ROC curve is useful when the biometric operates in the verification mode, since this curve expresses the quality of a 1:1 matcher. However, a biometric system can also operate in the identification mode, in which the quality of a 1:n matcher matters. For the assessment of this last mode, it is very useful to analyse the system performance through the Cumulative Match Characteristic (CMC) curve, since it allows to judge the ranking capabilities of an identification system (JAIN; ROSS; NANDAKUMAR, 2011).

5.3 Experiments

In this work, the experiments were carried out utilizing two sets of faces from the EURECOM Kinect database:

- Set 1: Gallery composed by the open mouth, smiling, and illumination variation face samples. Probe composed by the neutral face samples. This set was composed to assess the method performance in the presence of mild expression, and illumination variations;
- Set 2: Gallery composed by the open mouth, smiling, and illumination variation face samples. Probe composed by neutral and right side occluded face samples. This set was composed to assess the method performance in the presence of mild expression difference, illumination variation, and face occlusion.

5.3.1 Experiment 1

In this experiment it was utilized the Set 1 of faces.

The Figure 20 shows the CMC curves obtained by the proposed method (fusion of 3DLBP and HAOG), by the 3DLBP and HAOG methods without the fusion, and also by other methods proposed in the literature (Entropy and Salience Maps, FPLBP, HOG, and 3DPCA). The results

from the methods proposed in the literature are taken from (GOSWAMI et al., 2013). As one can see, the proposed method (fusion of 3DLBP and HAOG) presented the higher identification accuracy.

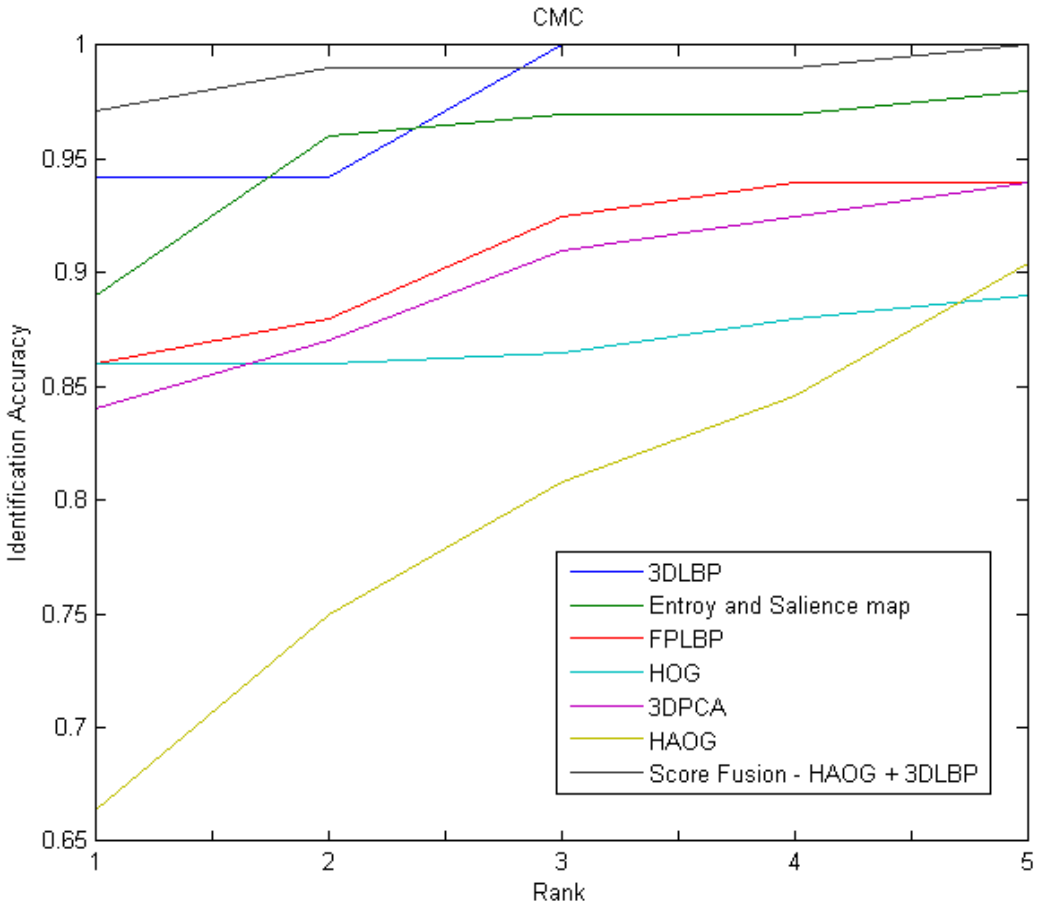


Figure 20 – CMC curves obtained for the 3DLBP, HAOG, 3DLBP fused with HAOG, Salience and Entropy map, FPLBP, HOG, 3DPCA.

In this experiment, the performance of the 3DLBP and HAOG methods, as well as their fusion, proposed in this work, were also assessed taking the two depth maps available: the one outputted directly by the Kinect, and that generated from the cloud points, after the symmetric filling process (as described in Section 4.4). The ROC curves obtained by using these two different depth maps are presented in Figure 21. Such curves show that, for the Set 1 of faces, the two depth maps obtained similar results.

Observing the ROC curve it is possible to see that the fusion between the 3DLBP and the HAOG with the original Kinect depth maps outperforms all the other methods. The 3DLBP with the original Kinect depth maps is the second better method. The fusion between HAOG and 3DBLP but with the depth maps generated from the cloud points has the third better performance.

It is important to notice that, even presenting the lower performance when executing alone, the HAOG descriptor contributes signicantly in order to increase the face recognition rates, when fused with 3DLBP method.

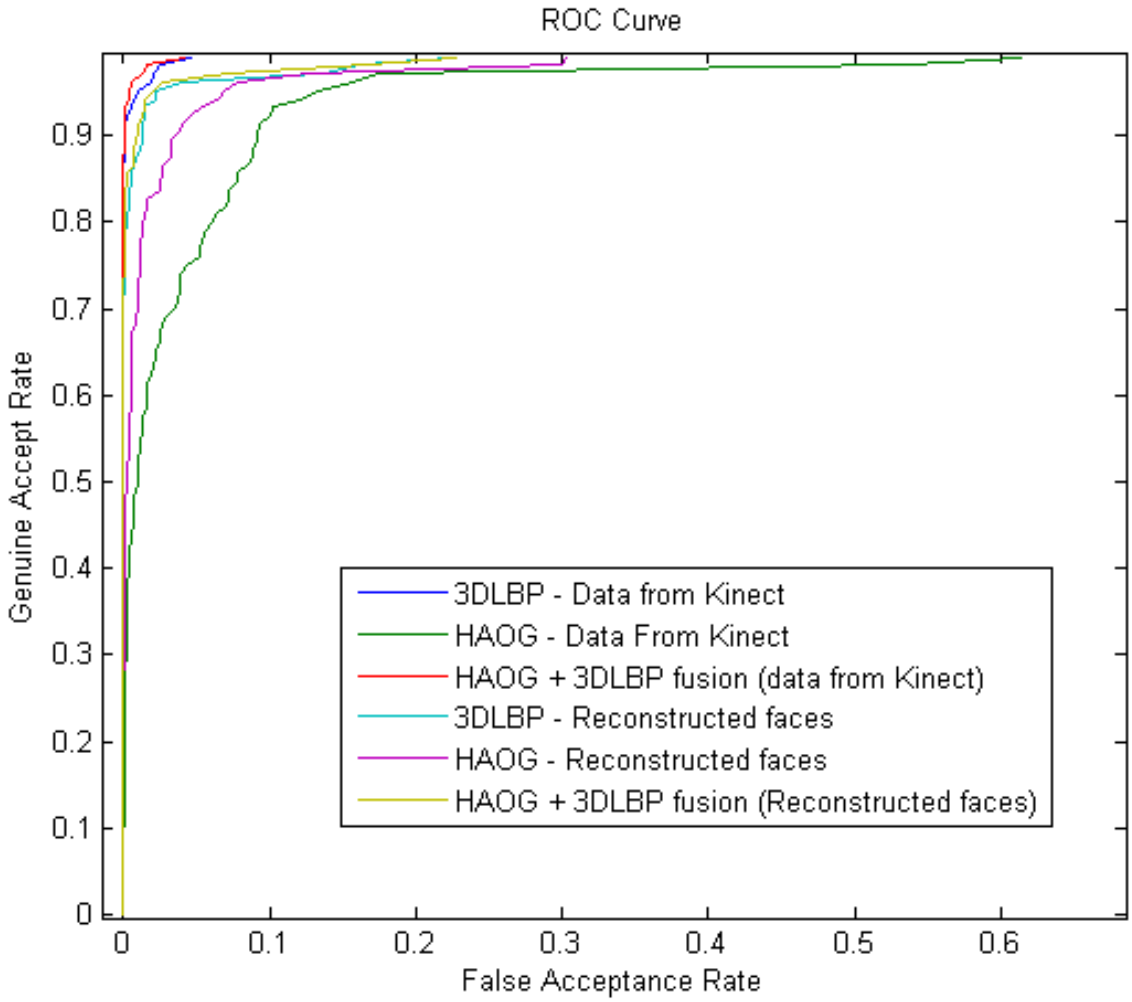


Figure 21 – ROC curves obtained for the 3DLBP, HAOG, and 3DLBP fused with HAOG, for the depth map generated directly by Kinect and the depth map obtained from the Kinect cloud points, using the Set 1.

The curves presented in Figures 20 and 21 were obtained with $w_1 = 0.8$ and $w_2 = 0.2$. Two CMC curves analyzing the fusion in this experiment were made, the Figure 22 presents the fusion utilizing the new depth maps and the Figure 23 utilizes only the original Kinect Depth map. Both cases, the initial value of w_1 was set to 0.1 and the initial value of w_2 was set to 0.9. Then, w_1 was incremented by 0.1 until reaches the value 0.9 and w_2 was decremented by 0.1 until reaches the value 0.1.

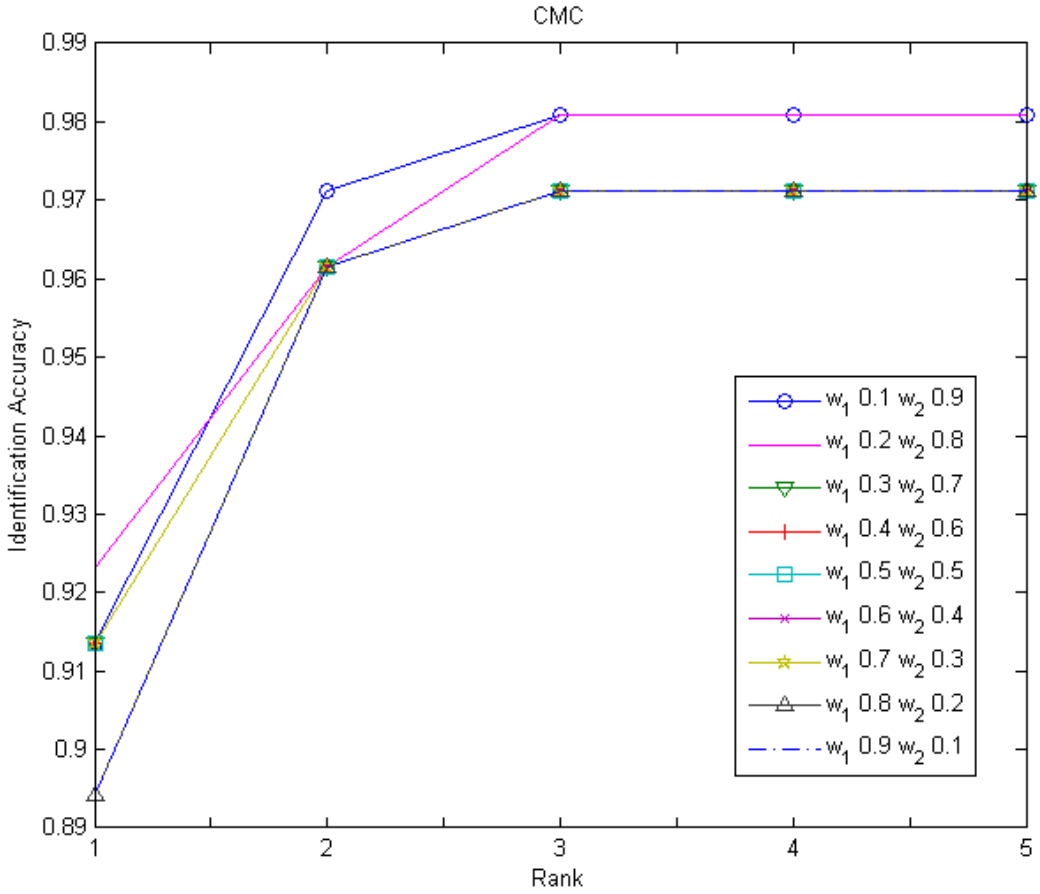


Figure 22 – CMC Curve comparing different values for w_1 and w_2, this fusion was made utilizing only the new depth maps.

With this experiment, it is possible to affirm that the score fusion for the 3DLBP and HAOG methods was effective in increasing the performance of the face recognition. The fusion proved to be effective even in the presence some level of facial expression variation.

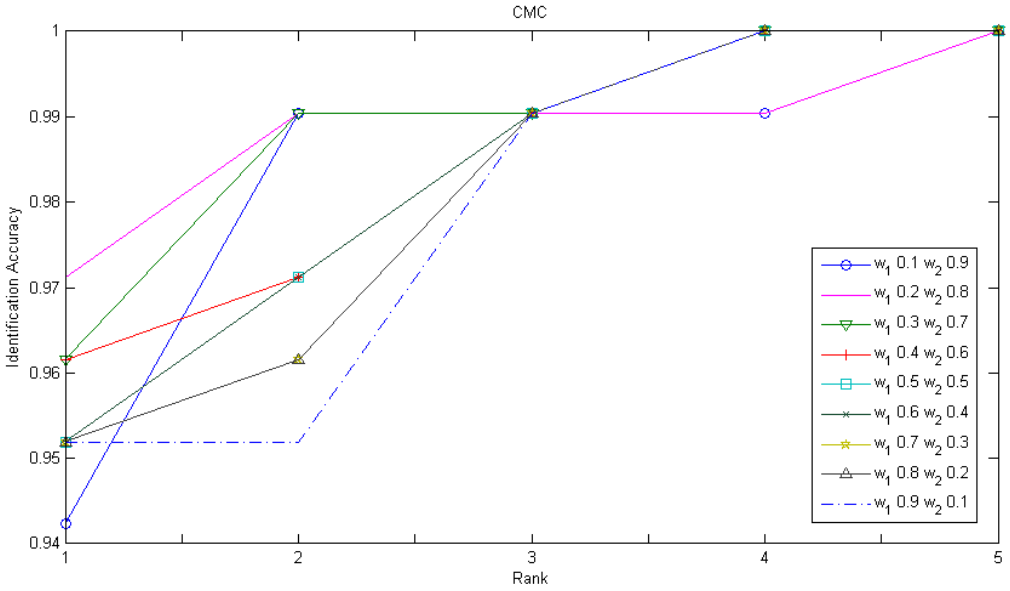


Figure 23 – CMC Curve comparing different values for w_1 and w_2, this fusion was made utilizing only the original Kinect depth maps.

In the Figure 23 it is possible to see the CMC Curve for the fusion on the original Kinect depth maps. Once more $w_1 = 0.8$ and $w_2 = 0.2$ has better performance on Rank 1 over other values.

Another important result that need to be highlighted is that the original Kinect images does not goes through heavy pre-processing. They are only cropped utilizing the landmarks annotated in the database.

5.3.2 Experiment 2

This experiment focused on the evaluation of the effects of the occlusion in the 3D face recognition. For this, two types of images are utilized; the original Kinect depth maps data and the new depth maps generated from the cloud points outputted by Kinect. The new maps were built as described in section 4.4.

In this experiment, the Set 2 of face images was utilized. In this set, the gallery is composed by the same faces presented in the Set 1, but the probe is composed also by the faces with their right side occluded. The Figure 24 shows the ROC curve with the 3DLBP, HAOG and their fusion applied in the original Kinect depth maps data and the same methods applied in the new depth maps generated from the cloud points data. HAOG with the original depth maps presented the worst performance, while the other methods with both types of depth maps data have similar results. Only the fusion of HAOG and 3DLBP with the new depth data highlights itself, it has a far better performance.

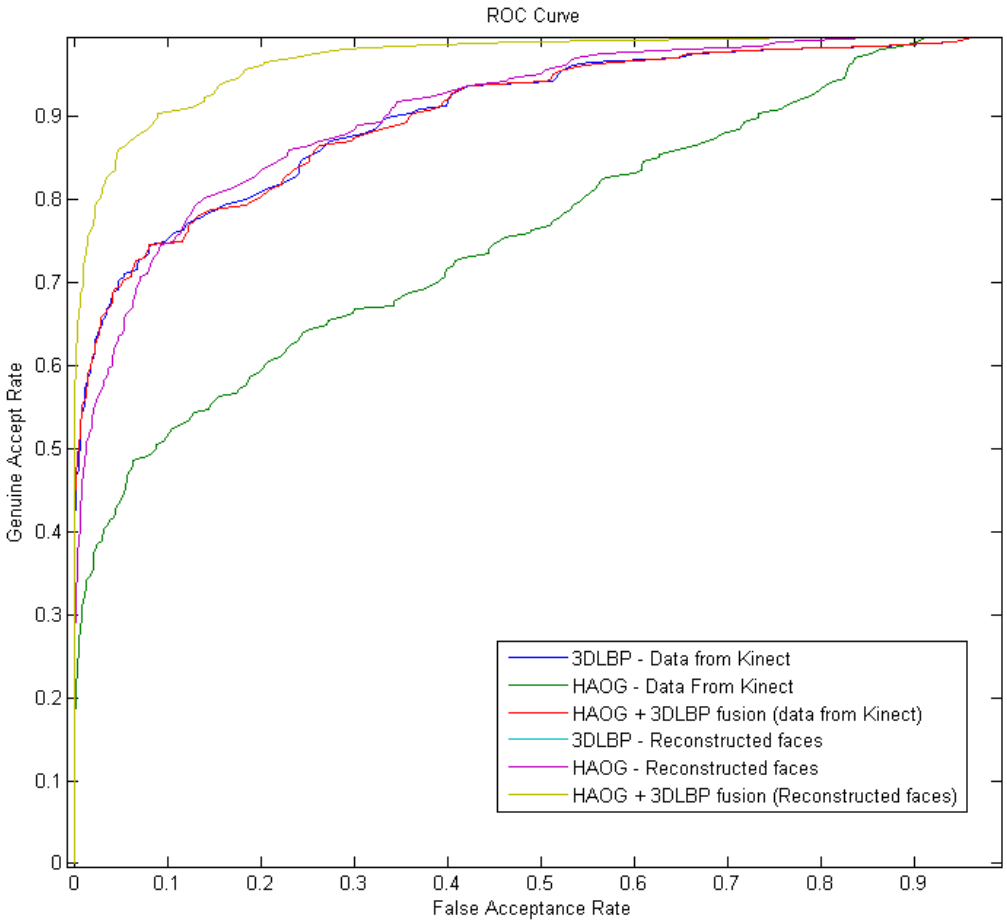


Figure 24 – ROC curves for the 3DLBP, HAOG, and the fusion between them. The results were obtained utilizing the Set 2 (which includes occlusion).

For the score fusion, this experiment utilizes the same values used in the experiment 1, that is $w_1 = 0.2$ and $w_2 = 0.8$.

The Figure 25 shows the CMC curves obtained with varying values of the weights w_1 and w_2. One can observe that in this experiment, even the performance for different weight values being closer to each other than in the first experiment, the pairs $(0.1, 0.9)$ and $(0.2, 0.8)$ for the weights (w_1, w_2) presented sligthly better results.

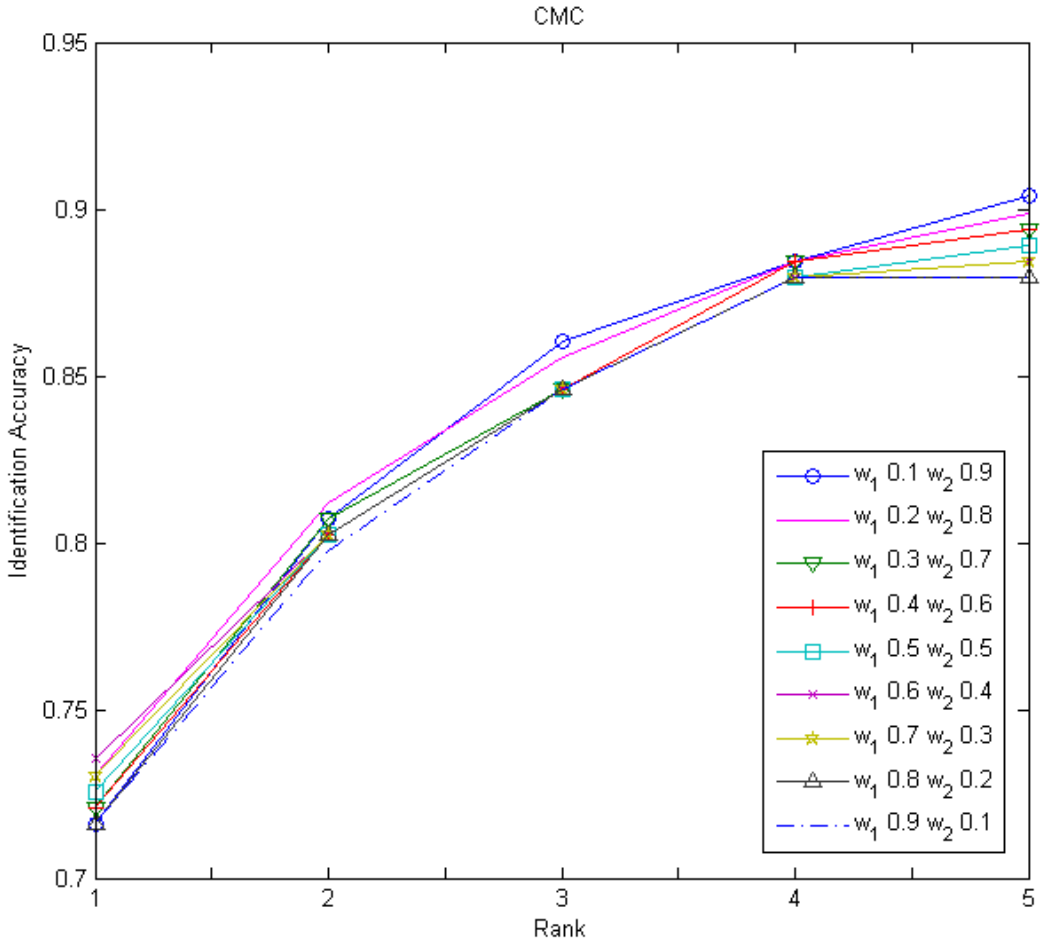


Figure 25 – CMC Curve comparing different values for w_1 and w_2. This experiments have an image with occluded face in the probe.

5.3.3 Experiment 3

Another 3D face recognition method that was assessed in our work is based on the work of Lei, Bennamoun & El-Sallam (2013). This method, described in Section 3.3.3, extract some characteristics (ACDN characteristics) directly from the point clouds, instead of from the depth maps.

In our implementation of this method, the ACDN characteristic are extracted only from the nose and eyes/forehead regions. Then, the histograms of all characteristics are concatenated and used as the feature vector that are submitted to a SVM classifier, in the matching stage.

Based on the methodology from the article proposed by Lei, Bennamoun & El-Sallam (2013), we have proposed an another characteristic, the P feature, which is the angle of the intersection from the line defined by the random point and the nose-tip and the Z axis.

Table 7 presents the results obtained with this experiment, which was carried out using the Set 1 of images, as in the previous experiments, and separately for each feature and the

fusion between them.

Table 7 – Recognition error rates for the experiment utilizing the ACDNP features on the face images from the Set 1.

Feature	Recognition Error %
A	80.77%
C	97.12%
D	94.23%
N	78.85%
P	83.65%
ACDNP	59.62%

Analysing the Table 7 it is possible to see that this method, originally proposed for high resolution 3D face data, does not perform well with Kinect data. The best result is the one obtained with the fusion of all the features (ACDNP) and, even in that scenario, the performance is only of 40.38% of correct recognition (42 subjects were correctly recognized from a set of 104).

This poor result is probably because of the low quantity of points from the cloud of points. Since the ACDN based methods depends on denser set of points, the data provided by the Kinect does not seems to be enough.

Therefore, the results obtained in our experiments lead us to conclude that for low resolution data, likewise the ones generated by Kinect, the facial features extracted from the depth maps (as the 3DLBP) can perform better than the facial features extracted directly from the cloud of points (as the ACDN).

Chapter 6

Conclusion

Throughout this work, it is possible to reach a few conclusions. It corroborates other works found in the literature showing that biometric identification can help to attenuate the problem with correct people identification.

Biometric features are not one hundred percent secure. Each biometric characteristic has different degrees of robustness against fraud. As can be seen in the Table 2 Face is very susceptible to fraud, especially when dealing with 2D images. The utilization of 3D data can help to increase the resistance to the aforementioned problem, since it is more difficult to spoof the face 3D information than a 2D image.

The major problem with 3D data is the high costs of the 3D scanners. There is also the problem related to the interaction of the subject with the sensor. With the traditional scanners, the subject needs to stay completely still during the scanning. Kinect can be a solution to these problems, but its data is sparse and of low resolution.

Another problem with Kinect is its maximum scanning depth (3.5 meters). High-resolution 2D cameras can caught several face images at a higher distance. This can be a problem depending on the type of the desired application.

From the experimental results obtained in this work and presented in Chapter 5, some conclusions can be pointed. The 3DLBP method, besides being effective for 3D face recognition using data captured with high resolution 3D sensors, can also be applied to 3D face recognition using data captured by low cost and low resolution sensors, like the Kinect. This is a very important result, since Kinect can be a viable alternative for the traditional 3D scanners that, besides being extremely expensive, in some cases can be very discomfortable to the users.

The results obtained in our work corroborates the results obtained by others, since we found that data generated by Kinect, even though being of lower resolution when compared with other traditional 3D sensors, are discriminative enough to allow a correct face recognition among different subjects.

In our work, only depth data is utilized. We do not use RGB data, since our focus was on the robustness to illumination changes.

Other important difficult in face recognition arises in situations with heavy face occlusion. In these cases, it is necessary to improve the face image quality. In this aspect, the Symmetric Filling approach, used in our method, proved to be of great importance, since it was able to reconstruct part of the face and increase the recognition rates of the proposed method.

The fusion of scores proved to be effective. In all cases the fusion obtained better results than utilizing only one method.

The results also showed that in real applications, it is more viable to utilize the depth map data generated from the cloud of points than the depth map outputted directly by Kinect, since the performance of the method is similar when the input data do not present face occlusions, but the performance can be greatly improved when the left or right side of the faces are occluded.

6.1 Contributions

At the end of this work, it is possible to highlight the following contributions:

1. A literature revision for 3D face recognition methods is presented;
2. A comparison of different techniques for 3D face recognition is presented;
3. A new method for 3D face recognition based on the fusion of 3DLBP and HAOG descriptor is proposed. This new method presents better results that the ones found in the literature. Besides, to the best of our knowledge, this was the first time that 3DLBP and HAOG features were proposed for 3D face recognition using Kinect;
4. A symmetric filling based approach is adopted to generate new depth maps from the cloud of points outputted by Kinect, in order to deal more robustly with occluded faces.

6.2 Future work

Suggestions for future work are:

- Explore more deeply the face occlusions, utilizing profile images;
- Performing real time face recognition;
- Utilizing deformation information to increase recognition performance in the presence of more intense facial expressions;
- Fuse information from the cloud point to increase performance recognition;
- Explore other types of fusion;
- Asses the proposed method using a database captured by the Kinect 2.0.

Bibliography

BESL, P. J.; MCKAY, N. D. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1992. IEEE Computer Society, Washington, DC, USA, v. 14, n. 2, p. 239–256, fev. 1992. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.121791>. Cited in page 27.

BOLLE, R.; PANKANTI, S. *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. Norwell, MA, USA: Kluwer Academic Publishers, 1998. ISBN 0792383451. Cited 2 times in pages 11 and 14.

CHEN, S. S.; DONOHO, D. L.; SAUNDERS, M. A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 1998. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, v. 20, n. 1, p. 33–61, dez. 1998. ISSN 1064-8275. Disponível em: <http://dx.doi.org/10.1137/S1064827596304010>. Cited in page 28.

DESIMONE, R.; DUNCAN, J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, 1995. v. 18, p. 193–222, 1995. Disponível em: <http://www.ncbi.nlm.nih.gov/sites-/entrez>. Cited in page 35.

ESTADÃO. *Autoescola é acusada de falsificar impressões digitais de alunos para vender CNH*. 2012. Disponível em: <http://www.estadao.com.br/noticias/cidades,autoescola-e-acusada-de-falsificar-impressoes-digitais-de-alunos-para-vender-cnh,924087,0.htm>. Acesso em: 28 out. 2012. Cited in page 15.

FAWCETT, T. An introduction to {ROC} analysis. *Pattern Recognition Letters*, 2006. v. 27, n. 8, p. 861 – 874, 2006. ISSN 0167-8655. {ROC} Analysis in Pattern Recognition. Disponível em: <http://www.sciencedirect.com/science/article/pii/S016786550500303X>. Cited in page 19.

GALOOGAHI, H. K.; SIM, T. Inter-modality face sketch recognition. *2012 IEEE International Conference on Multimedia and Expo*, 2012. IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 224–229, 2012. ISSN 1945-7871. Cited in page 41.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 2nd. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001. ISBN 0201180758. Cited in page 25.

GOSWAMI, G. et al. On rgb-d face recognition using kinect. *International Conference on Biometrics: Theory, Applications and Systems*, 2013. 2013. Cited 5 times in pages 22, 34, 35, 42, and 51.

HAYDAY, G. *2002 NTA Monitor Password Survey*. [S.l.], 2002. Cited in page 14.

HUANG, Y.; WANG, Y.; TAN, T. Combining statistics of geometrical and correlative features for 3d face recognition. In: *Proceedings of the British Machine Vision Conference*. [S.l.]: BMVA Press, 2006. p. 90.1–90.10. ISBN 1-901725-32-4. Doi:10.5244/C.20.90. Cited 3 times in pages 8, 39, and 40.

HUYNH, T.; MIN, R.; DUGELAY, J.-L. An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In: *ACCV 2012, Workshop on Computer Vision with Local Binary Pattern Variants, Daejeon, Korea, November 5-9, 2012 / Published also as LNCS, Vol 7728, PART 1*. Daejeon, KOREA, DEMOCRATIC PEOPLE'S REPUBLIC OF: [s.n.], 2012. Disponível em: <http://www.eurecom.fr/publication/3849>. Cited 2 times in pages 40 and 48.

ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998. IEEE Computer Society, Washington, DC, USA, v. 20, n. 11, p. 1254–1259, 1998. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.730558>. Cited in page 35.

JAIN, A. K.; MALTONI, D. *Handbook of Fingerprint Recognition*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003. ISBN 0387954317. Cited 3 times in pages 8, 15, and 16.

JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 2004. v. 14, p. 4–20, 2004. Cited 4 times in pages 9, 16, 17, and 44.

JAIN, A. K.; ROSS, A. A.; NANDAKUMAR, K. *Introduction to Biometrics*. [S.l.: s.n.], 2011. ISBN 0792383451. Cited 4 times in pages 8, 19, 20, and 50.

KAKADIARIS, I. A. et al. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007. IEEE Computer Society, Washington, DC, USA, v. 29, n. 4, p. 640–649, abr. 2007. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/TPAMI.2007.1017>. Cited 3 times in pages 12, 28, and 35.

KASS, M.; WITKIN, A. Analyzing oriented patterns. *Comput. Vision Graph. Image Process.*, 1987. Academic Press Professional, Inc., San Diego, CA, USA, v. 37, n. 3, p. 362–385, mar. 1987. ISSN 0734-189X. Disponível em: <http://dx.doi.org/10.1016/0734-189X(87)90043-0>. Cited in page 42.

LEI, Y.; BENNAMOUN, M.; EL-SALLAM, A. A. An efficient 3d face recognition approach based on the fusion of novel local low-level features. *Pattern Recognition*, 2013. v. 46, n. 1, p. 24–37, 2013. Cited 12 times in pages 8, 9, 22, 28, 29, 30, 31, 32, 35, 36, 37, and 56.

LEXISNEXIS. *2013 LexisNexis True Cost of Fraud Study*. 2013. Disponível em: <http://www.lexisnexis.com/risk/true-cost-fraud/>. Cited in page 11.

LI, B. et al. Using kinect for face recognition under varying poses, expressions, illumination and disguise. In: *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. [S.l.: s.n.], 2013. p. 186–192. ISSN 1550-5790. Cited 10 times in pages 9, 22, 23, 24, 27, 28, 36, 37, 38, and 42.

LI, S. Z.; JAIN, A. K. (Ed.). *Handbook of Face Recognition, 2nd Edition*. Springer, 2011. ISBN 978-0-85729-931-4. Disponível em: <http://dblp.uni-trier.de/db/books/daglib/0027896.html>. Cited 3 times in pages 11, 22, and 23.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 2004. Kluwer Academic Publishers, Hingham, MA, USA, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691. Disponível em: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>. Cited in page 24.

MARTIN, A. et al. The det curve in assessment of detection task performance. In: . [S.l.: s.n.], 1997. p. 1895–1898. Cited in page 19.

MIAN, A. Illumination invariant recognition and 3d reconstruction of faces using desktop optics. *Opt. Express*, 2011. OSA, v. 19, n. 8, p. 7491–7506, Apr 2011. Disponível em: <http://www.opticsexpress.org/abstract.cfm?URI=oe-19-8-7491>. Cited in page 27.

MIAN, A.; BENNAMOUN, M.; OWENS, R. Three-dimensional model-based object recognition and segmentation in cluttered scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2006. v. 28, n. 10, p. 1584–1601, 2006. ISSN 0162-8828. Cited in page 25.

MIAN, A.; BENNAMOUN, M.; OWENS, R. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007. IEEE Computer Society, Washington, DC, USA, v. 29, n. 11, p. 1927–1943, nov. 2007. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/TPAMI.2007.1105>. Cited 6 times in pages 22, 24, 25, 26, 35, and 36.

MIAN, A. S.; BENNAMOUN, M.; OWENS, R. A. A novel representation and feature matching algorithm for automatic pairwise registration of range images. *International Journal of Computer Vision*, 2006. v. 66, 2006. Cited in page 25.

MILLER, B. Vital signs of identity. *IEEE Spectr.*, 1994. IEEE Press, Piscataway, NJ, USA, v. 31, n. 2, p. 22–30, feb 1994. ISSN 0018-9235. Disponível em: <http://dx.doi.org/10.1109/6.259484>. Cited 3 times in pages 8, 14, and 15.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 1996. v. 29, n. 1, p. 51–59, 1996. Cited in page 38.

OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2002. v. 24, n. 7, p. 971–987, 2002. ISSN 0162-8828. Cited in page 38.

PHILLIPS, P. J. et al. Overview of the face recognition grand challenge. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. Washington, DC, USA: IEEE Computer Society, 2005. (CVPR '05), p. 947–954. ISBN 0-7695-2372-2. Disponível em: <http://dx.doi.org/10.1109/CVPR.2005.268>. Cited in page 27.

PLATT, J. C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: *ADVANCES IN LARGE MARGIN CLASSIFIERS*. [S.l.]: MIT Press, 1999. p. 61–74. Cited in page 33.

PRABHAKAR, S.; PANKANTI, S.; JAIN, A. Biometric recognition: Security and privacy concerns. *IEEE Security and Privacy*, 2003. IEEE Computer Society, Los Alamitos, CA, USA, v. 1, p. 33–42, 2003. ISSN 1540-7993. Cited 7 times in pages 8, 14, 15, 17, 18, 19, and 20.

RODRIGUEZ, Y.; MARCEL, S. Face authentication using adapted local binary pattern histograms. In: *9th European Conference on Computer Vision (ECCV)*. [S.l.: s.n.], 2006. IDIAP-RR 06-06. Cited in page 38.

SHPUNT, A.; ZALEVSKY, Z. *Depth-varying light fields for three dimensional sensing*. Google Patents, 2008. US Patent App. 11/724,068. Disponível em: <https://www.google.com/patents-/US20080106746>. Cited in page 23.

STOWERS, J.; HAYES, M.; BAINBRIDGE-SMITH, A. Altitude control of a quadrotor helicopter using depth map from microsoft kinect sensor. In: *Mechatronics (ICM), 2011 IEEE International Conference on*. [S.l.: s.n.], 2011. p. 358 –362. Cited in page 23.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 1994. v. 58, p. 267–288, 1994. Cited in page 28.

TIBSHIRANI, R. J. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 2013. v. 7, n. 0, p. 1456–1490, 2013. ISSN 1935-7524. Disponível em: <http://dx.doi.org/10.1214/13-ejs815>. Cited in page 28.

WAYMAN, J. Technical testing and evaluation of biometric identification devices. In: JAIN, A.; BOLLE, R.; PANKANTI, S. (Ed.). *Biometrics*. [S.l.]: Springer US, 2002. p. 345–368. ISBN 978-0-387-28539-9. Cited in page 18.

YANG, J.; LIU, C.; YANG, J.-y. What kind of color spaces is suitable for color face recognition? *Neurocomput.*, 2010. Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 73, n. 10-12, p. 2140–2146, jun. 2010. ISSN 0925-2312. Disponível em: <http://dx.doi.org/10.1016/j.neucom.2010.02.005>. Cited in page 28.

YIN, L. et al. A 3d facial expression database for facial behavior research. In: *Proc. IEEE Int'l Conf. Face and Gesture Recognition*. [S.l.: s.n.], 2006. p. 211–216. Cited in page 28.

YIP, A.; SINN, P. *Role of Color in Face Recognition*. 2001. Cited in page 28.

ZOLLHÖFER, M. et al. Automatic reconstruction of personalized avatars from 3d face scans. *Comput. Animat. Virtual Worlds*, 2011. John Wiley and Sons Ltd., Chichester, UK, v. 22, n. 2-3, p. 195–202, abr. 2011. ISSN 1546-4261. Disponível em: <http://dx.doi.org/10.1002/cav.405>. Cited in page 23.