University of Windsor
Scholarship at UWindsor

Electronic Theses and Dissertations | Theses, Dissertations, and Major Papers

5-11-2018

3D FACE RECOGNITION USING LOCAL FEATURE BASED METHODS

Sima Soltanpour
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation
Soltanpour, Sima, "3D FACE RECOGNITION USING LOCAL FEATURE BASED METHODS" (2018). *Electronic Theses and Dissertations*. 7416.
https://scholar.uwindsor.ca/etd/7416

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

3D FACE RECOGNITION USING LOCAL FEATURE BASED METHODS

by

Sima Soltanpour

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
at the University of Windsor

Windsor, Ontario, Canada

2018

© 2018 Sima Soltanpour

3D face recognition using local feature based methods

by

Sima Soltanpour

APPROVED BY:

_____

Y. L. Murphey, External Examiner
University of Michigan-Dearborn

_____

B. Boufama
School of Computer Science

_____

M. Ahmadi
Department of Electrical and Computer Engineering

_____

M. Khalid
Department of Electrical and Computer Engineering

_____

J. Wu, Advisor
Department of Electrical and Computer Engineering

April 27, 2018

Declaration of Co-Authorship / Previous Publication

I. Co-Authorship Declaration

I hereby declare that this dissertation incorporates material that is result of joint research, as follows:

This dissertation also incorporates the outcome of a joint research undertaken under the supervision of and in collaboration with Dr. Q. M. Jonathan Wu. The collaboration is covered through Chapters 2, 3, 4, and 5 of the dissertation. Chapter 2 of the dissertation was co-authored with Dr. B. Boufama, and chapter 3 was co-authored with M. Anvaripour. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of the collaborator was primarily through the provision of valuable suggestions for the representation of ideas, for the analysis of the results for the experiments carried out and editorial activities throughout the process of dissemination of the work.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my dissertation, and have obtained written permission from each of the co-authors to include the above materials in my dissertation.

I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

Thesis chapter	Publication title/full citation	Publication status
Chapter 2	S. Soltanpour, B. Boufama, and Q. M. Jonathan Wu, "A survey of local feature methods for 3D face recognition", Elsevier Pattern Recognition, 72, 391-406, 2017.	Published
Chapter 3	S. Soltanpour, Q. M. Jonathan Wu, and M. Anvaripour, "Multimodal 2D-3D face recognition using structural context and pyramidal shape index", IET International Conference on Imaging for Crime Prevention and Detection (ICDP), 2-6, 2015. S. Soltanpour, and Q. M. Jonathan Wu, "Multimodal 2D-3D face recognition using local descriptors: pyramidal shape map and structural context", IET Biometrics, 6(1), 27-35, 2016. (Invited paper from ICDP 2015)	Published Published
Chapter 4	S. Soltanpour, and Q. M. Jonathan Wu, "Multiscale depth local derivative pattern for sparse representation based 3D face recognition", IEEE International Conference on Systems, Man, and Cybernetics (SMC), 560- 565, 2017.	Published
Chapter 5	S. Soltanpour, and Q. M. Jonathan Wu, "High-order local normal derivative pattern (LNDP) for 3D face recognition", IEEE International Conference on Image Processing (ICIP), 2811-2815, 2017. S. Soltanpour, and Q. M. Jonathan Wu, "Weighted extreme sparse classifier and local normal derivative pattern for 3D face recognition", IEEE Transaction on Image Processing, 2018.	Published Submitted

II. Declaration of Previous Publication

This dissertation includes 6 original papers that have been previously published/submitted in peer reviewed conference proceedings and journals, as above table shows.

I certify that I have obtained a written permission from the copyright owners to include the above published materials in my dissertation. I certify that the above material describes

work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owners to include such materials in my dissertation.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

Face recognition has attracted many researchers' attention compared to other biometrics due to its non-intrusive and friendly nature. Although several methods for 2D face recognition have been proposed so far, there are still some challenges related to the 2D face including illumination, pose variation, and facial expression. In the last few decades, 3D face research area has become more interesting since shape and geometry information are used to handle challenges from 2D faces. Existing algorithms for face recognition are divided into three different categories: holistic feature-based, local feature-based, and hybrid methods. According to the literature, local features have shown better performance relative to holistic feature-based methods under expression and occlusion challenges.

In this dissertation, local feature-based methods for 3D face recognition have been studied and surveyed. In the survey, local methods are classified into three broad categories which consist of keypoint-based, curve-based, and local surface-based methods. Inspired by keypoint-based methods which are effective to handle partial occlusion, structural context descriptor on pyramidal shape maps and texture image has been proposed in a multimodal scheme. Score-level fusion is used to combine keypoints' matching score in both texture and shape modalities. The survey shows local surface-based methods are efficient to handle facial expression. Accordingly, a local derivative pattern is introduced to extract distinct features from depth map in this work. In addition, the local derivative pattern is applied on surface normals. Most 3D face recognition algorithms are focused to utilize the depth information to detect and extract features. Compared to depth maps, surface normals of each point can determine the facial surface orientation, which provides an efficient facial surface representation to extract distinct features for recognition task. An Extreme Learning Machine (ELM)-based auto-encoder is used to make the feature space more discriminative. Expression and occlusion robust analysis using the information from the normal maps are investigated by dividing the facial region into patches. A novel hy-

brid classifier is proposed to combine Sparse Representation Classifier (SRC) and ELM classifier in a weighted scheme.

The proposed algorithms have been evaluated on four widely used 3D face databases; FRGC, Bosphorus, Bu-3DFE, and 3D-TEC. The experimental results illustrate the effectiveness of the proposed approaches. The main contribution of this work lies in identification and analysis of effective local features and a classification method for improving 3D face recognition performance.

to my

dear parents

and loving husband Ashkan

Acknowledgements

I would like to thank my supervisor Dr. Q. M. Jonathan Wu for all his support during the entire program, patience, and belief in my abilities. I am also grateful to my doctoral committee members: Dr. Majid Ahmadi, Dr. Mohammed Khalid, and Dr. Boubakeur Boufama for their valuable suggestions and insightfulness. I thank the external examiner Dr. Yi Lu Murphey for her suggestions and help in improving the quality of the dissertation.

I would also like to thank the departmental graduate secretary, Ms. Andria Ballo for all her help during my studies at the University of Windsor. I am grateful to my friend, Dr. Ashirbani Saha for her help and support. I thank my friends and colleagues in the ECE department and CVSS laboratory whom have supported and believed in me.

I would like to express my gratitude and appreciation to my dear mother and father for instilling in me the love for learning and knowledge, and for their continuous support and encouragement not only for my studies but for my whole life. I would like to thank my sisters for their inspiration and love. I also thank my parents-in-law for their support and patience.

Finally, my deepest gratitude goes to my husband, Ashkan, for his tireless support. Without his unconditional love, help, and encouragement this research work would have never come to a successful completion.

Table of Contents

List of Tables

List of Figures

List of Acronyms

2D	Two Dimensional
3D	Three Dimensional
AIK	Adjustable Integral Kernels
AMD	Advanced Micro Device
ARS	Angular Radial Signatures
CMC	Cumulative Match Characteristic
CNPs	Closest Normal Points
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DoG	Difference of Gaussian
EER	Equal Error Rate
EID	Expression Insensitive Descriptor
ELM	Extreme Learning Machine
ESC	Extreme Sparse Classifier
FAR	False Acceptance Rate
FFT	Fast Fourier Transform
FRGC	Face Recognition Grand Challenge
FRR	False Rejection Rate

GA	Genetic Algorithm
GB	Giga Byte
GH	Geometric Histogram
GHz	Giga Hertz
HI	Histogram Intersection
HLNDP	Histogram of Local Normal Derivative Pattern
ICP	Iterative Closest Point
KMTS	Keypointbased Multiple Triangle Statistics
LBP	Local Binary Pattern
LCC	Local Coordinate Coding
LDP	Local Derivative Pattern
LNP	Local Normal Pattern
MB	Mega Byte
MsELBP	Multiscale Extended Local Binary Pattern
MsMcLNP	Multiscale Multicomponent Local Normal Pattern
MsLBP	Multiscale Local Binary Pattern
MsDLDP	Multiscale Depth Local Derivative Pattern
MsLNDP	Multiscale Local Normal Derivative Pattern
PC	Personal Computer
PCA	Principle Component Analysis

ROI	Region of Interest
ROC	Receiver Operating Characteristic
R3DM	Region based 3D Deformable Model
RR1	Rank-1 Recognition Rate
RAM	Random Access Memory
SHF	Spherical Harmonic Features
SI	Shape Index
SIFT	Scale Invariant Feature Transform
SRC	Sparse Representation Classifier
SSDM	Signed Shape Difference Map
SVM	Support Vector Machine
3DWWs	3D Weighted Walkthroughs
VGG	Visual Geometry Group
VLBP	Voting Local Binary Pattern
VR	Verification Rate
WESC	Weighted Extreme Sparse Classifier
WSRC	Weighted Sparse Representation Classifier

Chapter 1

Introduction

The work presented in this dissertation is related to the research and investigation for improving 3D human face recognition system performance using local features. In this chapter, the preliminary concepts related to the area of 3D face recognition are discussed. In addition, the motivation and objective of the proposed research are presented in this chapter.

1.1 Face recognition

Biometrics are physiological or behavioral characteristics of people measured and analyzed for the purpose of verifying identity. The extraction and representation of human characteristic have been an interesting research area in computer vision and pattern recognition for many years. Among biometrics, the human face attracts a lot of attention because of its applicability in important areas, such as security and surveillance. Compared to other types of biometrics such as iris images, finger-prints, palm-prints, and retinal scans; facial images are more socially acceptable since they are easily captured using contact-free scanners.

Face recognition is defined as a process to identify or verify a person's identity by comparing the input face characteristics against known faces from a database. A typical framework for a face recognition system has been shown in figure 1.1.

There are two different modalities including 2D images (grey scales and color images), and 3D data (depth maps, point clouds, and meshes) which are common for face recognition. The main focus of this research is on 3D face data because of the robustness under variations in lighting, head pose, and sensor viewpoint. 3D data can be captured using different 3D scanners via active or passive techniques [1]. Figure 1.2 illustrates some popular

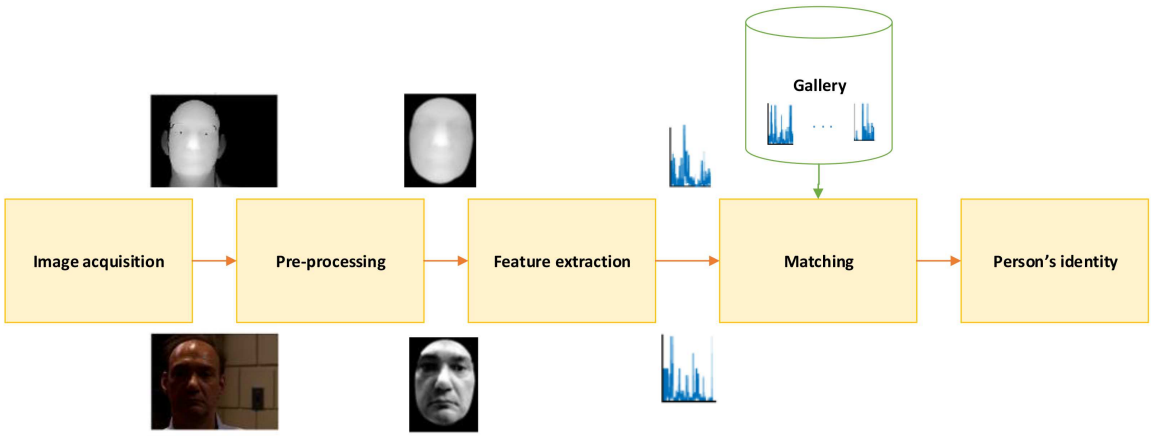

Figure 1.1 – A typical face recognition system framework

3D scanners. Before feature extraction and classification, any noise and spikes in captured data need to be removed. Noise is generated due to optical components of the sensors, the external ambiance, and the facial properties. Spikes are a common problem found in 3D captured data because of specular surfaces such as the eyes, the nose tip, and shiny teeth. In the pre-processing step, the noise and spikes are removed using filtering and thresholding techniques. The step of feature extraction is an approach to encode the distinct information of the face image. An efficient feature extractor should be discriminating for different subjects, compact, and robust under facial challenges. In the step of face matching, generally, two scenarios, which include identification and verification are performed. In the identification scenario, the identity of the input face is determined by searching the gallery to find the most similar face. During the verification scenario, the claimed identity of the input face is accepted or rejected by comparing the similarity between the probe face and the gallery face using a predefined threshold.

1.2 Applications

There are different applications of any face recognition system. It is used in game and movie industries for modeling and animating human characters. Facial expression is critical

Figure 1.2 – Popular 3D scanners [1]

to develop advanced human-machine interface [14]. In security and surveillance, it can be applied to a myriad of aspects such as system log on, internet access, access control, border control, suspect tracking, and terrorist identification. Medical treatments, such as facial surgery and maxillofacial rehabilitation have also emerged as other application based research directions for face recognition [1].

1.3 Challenges

1.3.1 Lighting

Lighting conditions for data acquisition can vary significantly for different subjects on various times/dates of capturing. Therefore, probe and gallery samples can be captured in different lighting conditions. Shadows and skin reflections can cause illumination variations in different samples. These changes cause an increase in intra-class variation which makes 2D face recognition a challenging task. Compared to 2D face data, 3D images are more robust under lighting variations due to shape and geometrical information.

1.3.2 Occlusion

Occlusion is one of the main problems in both 2D and 3D face recognition. It can occur due to the presence of glasses, caps, scarves, covering a part of the face by a hand, and hair (see some examples in figure 1.3).

Figure 1.3 – Examples of occlusion. From left to right, eye occlusion, mouth occlusion with a hand, occlusion caused by eyeglasses and hair [2]

Figure 1.4 – Some samples from happiness expression [2]

1.3.3 Facial expression

One of the major challenges for face recognition systems is expression variation. It can affect the performance of both 2D and 3D face recognition systems. The shape and geometry of the face can suffer deformations due to expression changes. Thus, facial expression variations can cause a significant difference between the samples of the same subject. Some samples from happiness expression have been presented in figure 1.4.

1.3.4 Pose

The probe and gallery samples can be captured with different poses. For example, one sample could be frontal, while, the other could contain a rotated face. Projective deformation

and self-occlusion have a remarkable influence on the accuracy of 2D face system. Since, pose correction methods can be applied for 3D face recognition, it is more robust under pose variations. However, extreme pose variation reduces the accuracy of 3D face system because of self-occlusion.

1.3.5 Generalization

Most 3D face recognition systems [15] have reported their performance on the Face Recognition Grand Challenge (FRGC) database [16] which is most used face database. Achieving high accuracy on one database cannot guarantee a good performance on other databases as each database consists of a subset of the challenges related to face recognition. Therefore, another important goal for a face recognition system is to achieve a good generalization performance.

1.4 Deep learning for face recognition

Recently, 2D face recognition systems have been improved by applying a deep convolutional neural network (CNN) [17] on public large-scale 2D face databases. Applying deep CNN for 3D face systems is not a straightforward task due to lack of large set databases. 3D scans are difficult to capture and the number of 3D samples and persons in public 3D face databases is limited. Kim et al. [18] presented deep 3D face recognition results. They reported the results on three public databases after fine-tuning the Visual Geometry Group, VGGFace network [17] on 3D depth images. An augmented database of around 100,000 depth images was used to tune the VGGFace network. Other databases are applied to the test phase individually. For all databases except one, their results do not outperform the state-of-the-art conventional methods. Moreover, they did not report the results on the most challenging 3D face database, FRGC [16], and their fine-tuned model is not publicly available.

1.5 Motivation

In the past decades, 2D face recognition has been comprehensively investigated [19, 20]. Although several methods have been proposed so far, there are still many limitations with 2D face recognition. 3D images compared to 2D data provide more reliable geometric information. Scale, rotation, and illumination do not affect the extraction of certain features, from these 3D images [21]. Furthermore, 3D pose estimation is more accurate than pose estimation with 2D images. Given these advantages, 3D face recognition has become an active research area aimed at overcoming the existing challenges, arising from 2D face images. A 3D face contains geometrical information and can be applied to overcome the challenges arising from illumination and pose. 3D face recognition under facial expression, extreme pose, and occlusion is still a very challenging task due to large intra-class variations.

Face recognition approaches are categorized based on the type of data and the algorithm used. This data is divided into three types: (i) 2D texture image (2D face recognition), (ii) 3D depth map or point clouds (3D face recognition), and (iii) 2D and 3D face data (multimodal face recognition) [22]. Approaches for 2D and 3D face recognition are divided into three categories, including (i) holistic feature-based, (ii) local feature-based, and (iii) hybrid algorithms [19]. Local features have attracted many researchers' attention due to their robustness under pose variation, expression, and occlusion [23]. Therefore, the main focus of this research is on 3D face recognition using local features.

1.6 Main Contributions

The major contributions of the dissertation are listed as follows:

- As an initial step in the field of 3D face recognition, the existing algorithms in three different categories, local feature-based, global feature-based, and hybrid methods have been studied. It was found that among these three categories, the local features

are more robust and promising to enhance the performance of the recognition system. Consequently, a complete survey was performed on the local feature-based methods for 3D face recognition.

- A methodology is proposed for hybrid face recognition using local features from shape and texture modalities incorporating histogram matching. This approach achieves improved performance on two databases as compared to the state-of-the-art algorithms. The overall performance of the proposed hybrid method is also better than state-of-the-art algorithms.

- An algorithm is proposed to extract local patterns on depth maps and surface normals. The proposed local descriptor improves the performance of the state-the-art algorithms on several databases.

- A methodology is proposed using the combination of two different classifiers to enhance the performance of the recognition system in terms of accuracy and computational cost. The approach achieves competitive performance compared to the state-of-the-art techniques in several databases.

1.7 Organization

This dissertation is organized into six chapters as follows:

- **Chapter 2** surveys the local feature-based methods for 3D face recognition in three different categories including keypoint-based methods, curve-based methods, and local surface-based methods. The most common 3D face databases are reviewed and the criteria to evaluate the recognition system is explained. The comparison between state-of-the-art approaches is discussed with their relative strengths and weaknesses.

- **Chapter 3** proposes a novel approach by the combination of 2D and 3D face data to improve the recognition performance. A new local descriptor is extracted on dif-

ferent facial shape maps and the texture image. A fusion scheme is used for hybrid matching and the final decision.

- **Chapter 4** introduces the Multiscale Depth Local Derivative Pattern (MsDLDP) descriptor to extract efficient local features from depth maps. Details of how the expression variation problem is addressed by excluding non-rigid areas of the face and applying Sparse Representation Classifier (SRC) are discussed.

- **Chapter 5** proposes a Weighted Extreme Sparse Classifier (WESC) to handle facial expression and occlusion using (Local Derivative Pattern) LDP on surface normals. The details of the weighted hybrid classifier and local pattern on the normal maps are described.

- **Chapter 6** summarizes the research findings and concludes this dissertation by indicating the scope of the future work.

Chapter 2

Literature Survey

One of the main modules in a face recognition system is feature extraction, which has a significant effect on the whole system performance. In the past decades, various types of feature extractors and descriptors have been proposed for 3D face recognition. Although several literature reviews have been carried out on 3D face recognition algorithms, only a few studies have been performed on feature extraction methods. The latter have a vital role to overcome degradation conditions, such as face expression variations and occlusions. Depending on the types of features used in 3D face recognition, these methods can be divided into two categories: global and local feature-based methods. Local feature-based methods have been effectively applied in the literature, as they are more robust to occlusions and missing data. This survey presents a state-of-the-art for 3D face recognition using local features, with the main focus being the extraction of these features.

2.1 Introduction

A number of surveys have been published in 3D face recognition during the last decade. Most of the earlier surveys have focused on the introduction, general summarization, and challenges of face recognition algorithms [24, 25, 26, 27]. A survey by Scheenstra et al. [25] reviewed 3D face recognition approaches in four different categories, and compared them with 2D face recognition methods. 3D face recognition methods alone or in combination with 2D intensity images were discussed in [26]. Various challenges for 2D and 3D face recognition were addressed and the limitations and solutions for different methods were discussed in [27]. Smeets et al. [28] conducted a survey on 3D face recognition by

summarizing the main characteristics and challenges of these approaches. A recent survey by Zhou et al. [29] covered different algorithms by categorizing them into single-modal and multimodal approaches, along with their advantages and disadvantages. Some of the recent review papers have focused on a specific challenge in face recognition. For instance, a survey on pose-invariant face recognition approaches is presented in [30], a comparative study on 3D face methods, under facial expression challenges, can be found in [31], and [32, 33] represent a survey on 3D facial expression recognition. In [1], only feature extraction and selection methods were investigated, for both 2D and 3D face recognition. The main focus of [1] was the presentation of the different methods, with less emphasis on the comparison of their advantages and drawbacks.

Approaches for 3D face recognition can be divided into three broad categories: holistic, feature-based, and hybrid matching methods [19]. In holistic matching methods, the focus is on the global similarity of faces. The entire 3D face (or model) is described by defining a set of global features. Examples in this category include the principle component analysis (PCA)-based method [34], the deformation modeling [35], the signed shape difference map (SSDM) [36], spherical harmonic features (SHF) [37], closest normal points (CNPs) [38], and region based 3D deformable model (R3DM) [39]. Feature-based matching methods rely on finding similar local features from the face or from special regions of the face (e.g., eyes and nose). Hybrid approaches are defined based on the combination of different types of approaches (holistic and feature-based) or data (2D and 3D images).

There are several reasons that make local methods more promising than holistic ones. In particular, in local methods complete models are not necessary and occlusions can be easily handled [40]. According to the survey by Abate et al. [27], applying local features is one possible solution for recognizing partially occluded faces. Recent survey by Zhou et al [29] mentioned different challenges for face recognition are pose, viewpoint, and expression that feature-based methods address these problems. Moreover, local descriptors, like Scale Invariant Feature Transform [41] and Local Binary Pattern [42], have yielded remarkable results in 2D face recognition. Because the main focus of local descriptors is on the shape

details, global or holistic methods perform better in similarity search applications, while local methods are more suitable for matching, identification and verification [40]. Mian et al. [3] mentioned one limitation for holistic methods : they need accurate normalization for pose and scale. Generally, the recognition performance of global features is usually affected by pose and/or scale variations. To solve this problem, manual and automatic landmark detection are used for normalization, with the manual one being more accurate. However, it makes the whole process semi-automatic, as in [43]. Recently, Gilani et al. [39] proposed a landmark detection technique for holistic methods. It uses a deep landmark identification network and needs a training step with synthetic images. Although, holistic algorithms apply all the visible facial shape information to create discrimination, obtaining the needed accurate pose normalization is not easy under noisy or low-resolution 3D scans. In this case, local features may perform better [44]. Furthermore, local methods can be robust under facial expressions, because sensitive facial regions can be excluded [31, 3]. In particular, local features can be extracted from the rigid parts of the face that are the least influenced by expression changes [13].

Based on the above discussion, local feature-based methods are a promising research topic for 3D face recognition application. We have conducted a survey on local methods to cover recent works in this area. In particular, none of these surveys specifically focuses on local feature-based 3D face recognition. Unlike [1] that presents feature extraction algorithms for both 2D and 3D, local and holistic features in combination with feature selection and fusion techniques, the main focus of this survey is a comprehensive study and comparison of different local feature-based techniques for 3D face recognition only. Compared to [25] that discusses local and global features, our survey covers more recent local-based works with more details on their performance, under different facial challenges. Therefore, this chapter provides a survey on various categories of local 3D features, together with comparisons as well as their limitations and advantages. The survey also aims at helping researchers to get a good overview on 3D face recognition, and enable them to select the most effective method for the right situation.

The remainder of this chapter is organized as follows. The terminology of 3D face recognition and databases are described in Section 2. Section 3 provides a comprehensive survey of local feature-based methods for 3D face recognition, including methods categorization and a detailed review of feature extraction algorithms. Section 4 presents a discussion on the reviewed methods and their comparison, while Section 5 concludes this survey, presenting potential future research directions.

2.2 Terminology and 3D databases

There are two scenarios for a typical 3D face recognition system; namely verification (1:1 matching) and identification (1:N matching). For identification, an unknown face (probe) is matched against known individuals (gallery) to find the best match. Verification refers to the confirmation or rejection of a claimed identity of a probe face. Furthermore, usually two metrics are considered for measuring the performance of a face recognition system. The Receiver Operating Characteristic (ROC) curve is used to measure the verification accuracy. ROC plots the False Rejection Rate (FRR) or Verification Rate (VR) against the False Acceptance Rate (FAR); at various thresholds, and interpolates between these points. FRR refers to the probability of incorrectly rejecting a person (two samples belonging to the same person) and FAR refers to the probability of accepting an incorrect person (two samples from two different people). The Cumulative Match Characteristic (CMC) curve, used to evaluate identification performance, plots the recognition rate against a number of ranks. The same matching threshold is used for both verification and identification scenarios.

2.2.1 Databases

There are different types of 3D data applied in the recognition system. Polygonal meshes of 3D faces are usually used in 3D face recognition applications for computational efficiency. Other types of 3D data include point clouds, a collection of 3D point coordinates,

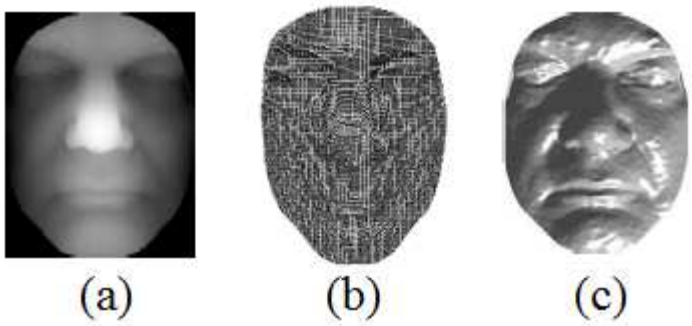

Figure 2.1 – 3D face data representations : (a) Range image, (b) Poin cloud, (c) Mesh [1]

and range images or depth maps, where each element represents the distance of a point from the sensor or from another reference point. Figure 2.1 illustrates a range image, point cloud and mesh representation as different types of 3D data. There are two types of acquisition systems for capturing 3D faces: active, like laser scanners and structured light, and passive, like stereo-based systems [1]. In active capturing systems, such as Minolta vivid scanners, triangulation technique is used. A laser line is shined on the face from a scanner and an image of the line is recorded by a camera. Although the accuracy of this method for 3D face acquisition is relatively high, it is time consuming. In structured light, for example Inspeck Mega Capturor II 3D, a pattern of light is projected on a face from a light source and the deformations of the pattern are measured using a camera. This technique is fast, but the captured data contains a number of holes and artifacts. In passive techniques based on stereo systems, for instance 3DMD digitizer and Di3D, two cameras are employed to capture the location of each point by matching corresponding pixels in two images. Because of the difficult and time-consuming problem of dense pixel matching, due to the relative uniformity of a human face for two images, the accuracy of this system is comparatively low [45]. To evaluate 3D face recognition algorithms, many databases have been created. Table 2.1 describes the currently popular 3D face databases in four different categories, according to 3D data type, and provides some details for each.

Table 2.1 – Popular 3D face databases

Reference/Name	Data Type	Intensity image	Number of subjects	Number of images	Scanner
[46]/FSU	mesh	no	37	222	Minolta Vivid 700
[47]/GavabDB	mesh	no	61	549	Minolta Vi-700 laser range scanner
[48]/FRAV3D	mesh	yes	105	-	Minolta Vivid 700 red laser light scanner
[49]/BU-3DFE	mesh	yes	100	2500	Stereo photography, 3DMD digitizer
[50]/UoY	mesh	yes	350	5000	Stereo vision 3D camera
[16]/FRGCv1	range image	yes	273	943	Minolta Vivid 3D scanner
[16]/FRGCv2	range image	yes	466	4007	Minolta Vivid 3D scanner
[51]/UND	range image	yes	277	953	Minolta Vivid 900
[52]/CASIA	range image	no	123	4059	Minolta Vivid 910
[53]/ND2006	range image	yes	888	13,450	Minolta Vivid 910
[35]/MSU	range image	no	90	533	Minolta Vivid 910
[54]/SHREC08	range image	no	61	427	-
[55]/3D-TEC	range image	yes	214	428	Minolta scanner
[56]/SHREC11	range image	no	130	780	Escan laser scanner
[57]/UMB-DB	range image	yes	143	1473	Minolta Vivid 900 laser scanner
[58]/Texas 3DFRD	range image	yes	118	1149	MU-2 stereo imaging system
[2]/Bosphorus	point cloud	yes	105	4666	The Inspeck Mega Capturor II 3D scanner
[59]/BU-4DFE	3D video	yes	101	60600	Di3D (Dimensional Imaging) dynamic system

2.3 3D local feature-based methods

In the context of face recognition, 3D local feature descriptors are built from 3D local facial information. These features have some advantages over global features, as global descriptors are more sensitive to pose, facial expressions and occlusions [44]. The main objective of local feature extraction methods is the detection of distinctive compact features, that are robust to a set of nuisances. To the best of our knowledge, local feature-based algorithms can be more robust against facial variations such as expression and occlusion, as they exclude parts that might be affected by those changes. In particular, there is no set of local

attributes that are completely invariant under all variations [26].

A number of 3D local feature descriptors for 3D face recognition have been presented in the literature. This section surveys and explains the main existing 3D local descriptors and groups them into three different categories: Keypoints-based, curve-based and local surface-based methods.

2.3.1 Keypoints-based methods

3D keypoints are interest points of shape, based on the definition of saliency. They are detected according to some geometric information of the surface. The methods typically involve two major steps, keypoint detection and feature description [60]. Although these methods can cope with occlusions and missing parts, their computational cost is much higher as they use a large number of keypoints, described by high dimensional feature vectors. Hence, it is very important to only select the most effective keypoints, from the local descriptors, to create an efficient feature vector.

Methods based on SIFT-like keypoints

Scale invariant feature transform (SIFT) [41] is a successful keypoint detector that has motivated researchers to use the same scheme in the case of 3D images. The most important limitation of SIFT keypoint-based methods is their sensitivity to noisy data. However, these methods do not require very sophisticated registration algorithms. Furthermore, the convincing representation of SIFT features on shape maps motivated researchers to apply this framework in 3D.

A framework to detect SIFT-inspired 3D keypoints was first proposed by Mian et al. [3], where they use the shape variation in combination with 2D SIFT descriptors. To detect 3D keypoints, for points in the sphere of radius r and center p, the mean vector m and covariance matrix C are calculated. Then, matrix V of the eigenvectors is obtained by performing principal component analysis (PCA) on C. A point p is defined as a keypoint,

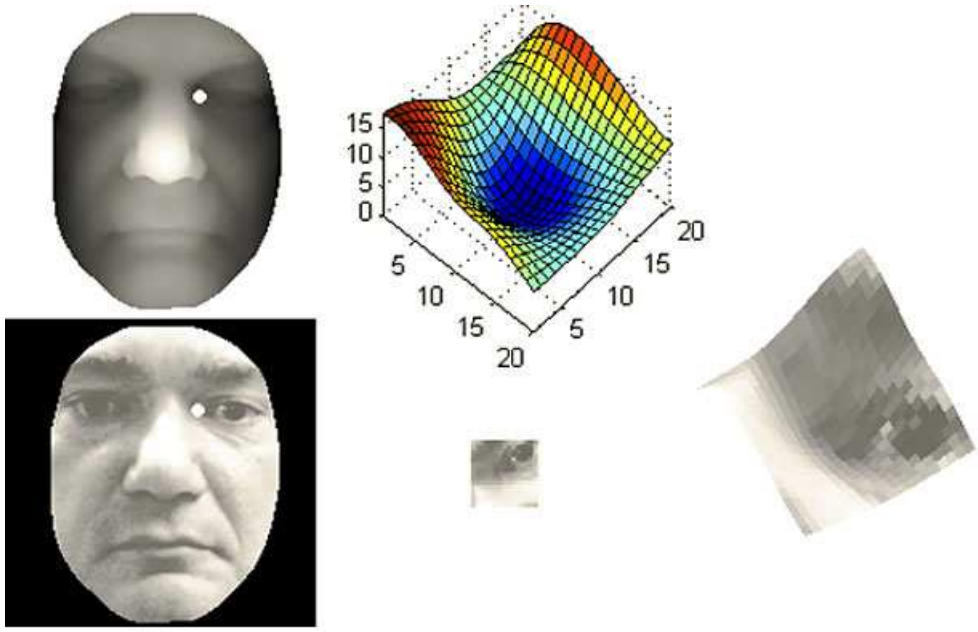

Figure 2.2 – A keypoint on a 3D face and its corresponding texture [3]

if the difference between the first two principle axes of the local region is greater than a threshold t. Figure 2.2 illustrates a keypoint on the 3D face and its corresponding texture image. This method has influenced other researchers; for example SIFT keypoints are used in [61] to detect relevant interest points on depth images, then local shape descriptors are defined for the neighborhood of each keypoint. Mayo and Zhang [62] proposed a multiview keypoint matching method, where SIFT keypoints are extracted from 2.5D images. In [63], SIFT descriptors are extracted from 2D matrices of curvature maps, where the features are defined at fixed scales and orientations for fixed locations. SIFT keypoint detection is applied on multiscale local binary pattern and shape index maps in [64], and on pyramidal shape index map in [65] for 3D domain and in combination with 2D keypoints, respectively. The extension work of [65] has been presented in [66] using curvature maps. The main weakness of these methods is their sensitivity to pose variations. Recently, a Keypoint-based Multiple Triangle Statistics (KMTS) method has been presented by Lei et al. [67] to handle pose variations where 3D keypoints are detected based on the method in [3].

Low-level geometric features [13], described in Section 3.3 of this chapter, are extracted from the patch around the detected keypoints. Applying low-level geometrical features without any complicated mathematical operation shows that the approach is time efficient. According to the experiments reported by authors using an Intel Core 2 Quad CPU and 16 GB RAM, the pre-processing, feature extraction and identification takes 0.62 s, 5.46 s, and 1.82 s, respectively.

Mesh-based methods

Although SIFT-like detectors present the informative features without registration for nearly frontal scans, they are sensitive to large pose variations or occlusions. To overcome these limitations, SIFT keypoints detection is applied directly on 3D mesh data in recent works. Generally, approaches using 2D keypoints detection ideas have allowed the discovery and implementation of powerful keypoint detectors in the 3D domain.

An extension of SIFT for 3D meshes, called MeshSIFT, was proposed by Maes et al. [68], then extended by Smeets et al. [4]. The approach consists of four major steps: keypoint detection, orientation assignment, local feature description, and feature matching. Given an input mesh M, the mean curvature H is calculated for each vertex i, at each scale s, to detect salient points. The normal vector of the keypoint neighboring vertices is projected onto its tangent plane. A weighted histogram is constructed using the projected normal vectors. The canonical orientations are estimated with the highest peak in the histogram. Normals and their projections onto the tangent plane are illustrated in Figure 2.3a. A feature descriptor is defined for each keypoint by the concatenated histograms of nine circular regions (the shape index and angles between normals), as shown in Figure 2.3b.

The meshSIFT-like keypoint detector has also been applied in [5] using maximum (k_{max}) and minimum (k_{min}) curvatures, estimated in the 3D Gaussian scale space. A salient point is the vertex whose value is a local extrema within its neighborhood. The detection of keypoints is illustrated in Figure 2.4a. To calculate the local descriptor, a geodesic disk with radius R is considered around each keypoint. Then, a circle with radius r_1 and eight

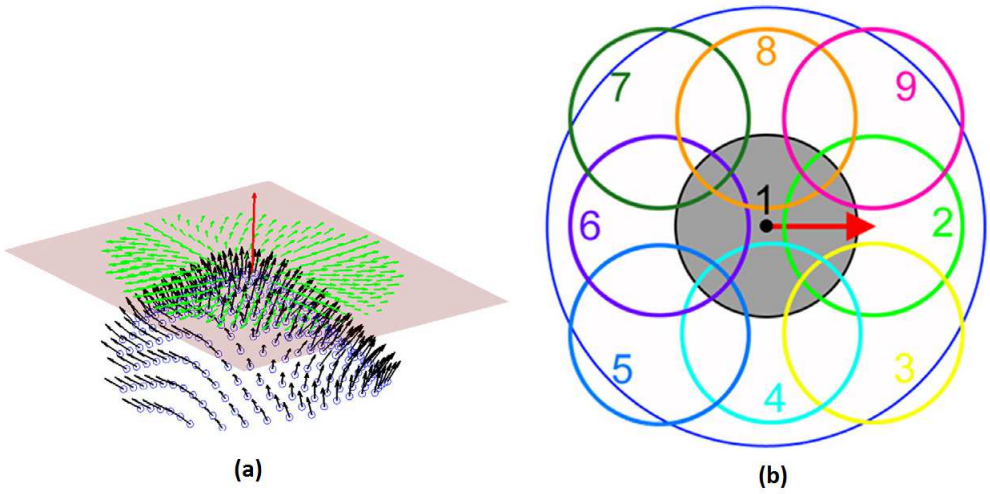

Figure 2.3 – a) Normals and their projections, b) Nine circular regions around a keypoint[4]

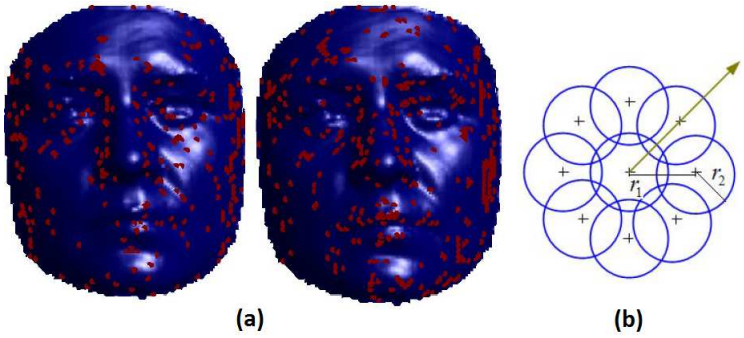

Figure 2.4 – a) Salient points by k_{max} (left) and k_{min} (right), b) Canonical orientation, salient point and its neighborhood vertices [5]

circles with radius r_2 are extracted, as shown in Figure 2.4b. Three histograms, including surface gradient (HoG), shape index (HoS), and gradient of shape index (HoGS) are calculated for each circle. The concatenation of these three histograms is considered the local descriptor. A common disadvantage of the above methods is the detection of a large number of keypoints. None of them presented a solution for selecting salient keypoints.

To overcome the stated problem and extract repeatable keypoints on 3D meshes, Mesh-DoG [6], an extension framework of [69], proposes a multiring geometric histogram (GH) as a descriptor. Given a 3D mesh, the mean curvature at each vertex is first computed. The

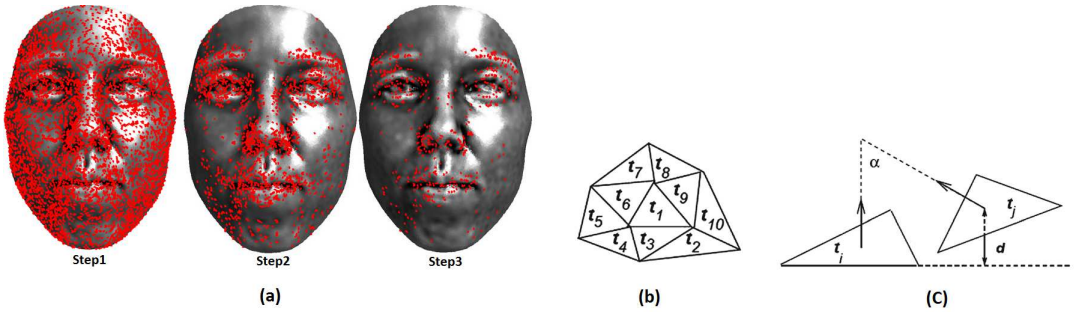

Figure 2.5 – a) 3D keypoints detected at each step b) central facet t_1 and its neighbors c) the angle and perpendicular distance [6]

detection of the 3D keypoints is done in three steps, i.e., scale-space, percentage threshold, and corner analysis. Figure 2.5a shows the detected 3D keypoints in the three different steps and GH computation. The normals and the difference between minimum and maximum perpendicular distance of two facets are calculated to create descriptors (see figure 2.5). An extension of the framework in [5] is described in [70], where a fine-grained matching of 3D keypoint descriptors has been proposed to handle degradation conditions. Among the above mentioned mesh-based SIFT-like matching methods [4, 70] provide a registration-free recognition scheme.

Recently, Elaiwat et al. [7] proposed a keypoint detector and a local feature descriptor by integrating different Curvelet elements of different orientations. Since Curvelet transform is based on FFT, the computational complexity of keypoint detection and descriptor definition is lower than SIFT-based methods. The coefficients of these Curvelet elements are computed at each scale a and angle θ, as shown in figure 2.6. Keypoints are detected by comparing the magnitudes of Curvelet coefficients with the mean value of all coefficients at scale a. A local descriptor is defined around each keypoint in the Curvelet domain for all the sub-bands of the scale in which the keypoints are extracted. Since the keypoints are extracted on different frequency bands and directions, they are highly repeatable and informative.

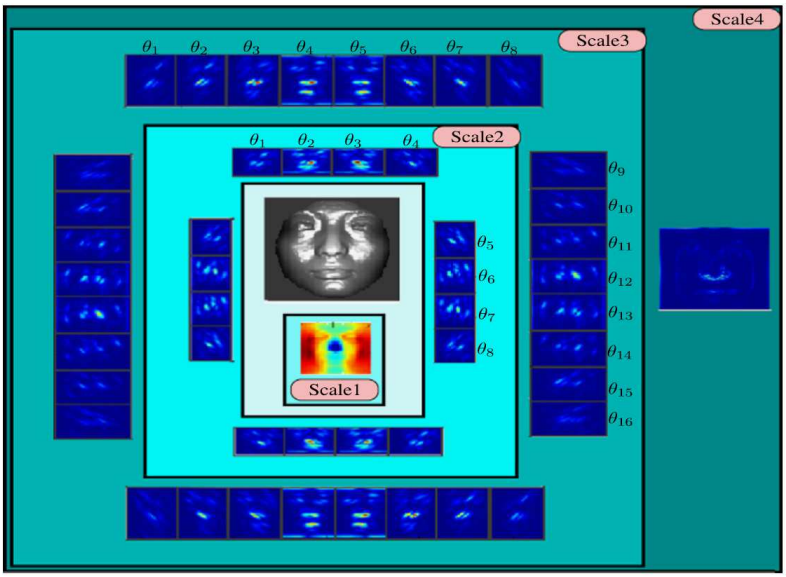

Figure 2.6 – Four-scale Curvelet decomposition [7]

Landmark-based methods

Landmarks are facial points extracted according to anatomical studies of the face. Some methods use a set of landmarks on the face to create feature vectors, obtained by calculating relationships between these landmarks. Therefore, accurate extraction of these landmarks is critical to generate reliable local features. The landmarks could be the eyes, nose and mouth on the facial image. They are also employed to correct the pose in pose sensitive local feature-based methods. Their disadvantage is the sparsity that can affect the recognition performance.

Shape index as curvature map is widely used to detect landmarks. In [71], feature points, included inside and outside the corners of the eyes and the nose tip, are extracted by calculating the local shape index at each point of the 3D mesh. In [72], shape index and spin images are used as local descriptors to extract landmark points. Spin image encodes each point p on the 3D face surface, with respect to the normal vector n at that point. Facial curvatures are also used for landmark detection [73, 74]. Triangles, resulting from the connection of the detected eyes and nose, are used in the recognition stage in [73]. In [74],

14 manually detected landmarks are used to define a local shape dictionary, consisting of curvature maps. Keypoints are then extracted from this shape dictionary. An enhanced version of this work, presented in [21], uses a non-linear machine learning approach, namely AdaBoost, to detect keypoints. There are other proposed methods to detect landmarks. In [75], five landmarks on a range image of the face are extracted using radial symmetry and shape information. These facial feature points are employed to extract a very small subset of points on probe images, that are invariant under facial expressions. Gupta et al. [76] presented an anthropometric approach by detecting 10 fiducial points and calculating the Euclidean and geodesic distance between them as features. Song et al. proposed a landmark localization approach that uses local coordinate coding (LCC) and consists of two stages: nose detection and resampling [77]. Another landmark-based method is described in [78], where the authors proposed an automatic 3D landmark localization method that can handle missing parts, with asymmetry pattern and shape regression. Recently, an automatic 3D facial landmark detection has been proposed in [79] using 2D Gabor wavelet features.

Summary

Table 2.2 summarizes the keypoint-based methods. The latter are categorized into SIFT-like, mesh-based, and landmarks. The neighborhood of a keypoint is defined based on three different measurements [80], i.e., Euclidean distance, geodesic distance and multirings. Methods based on geodesic distances are robust under isometric deformations. On the other hand, the geodesic distance calculation is time-consuming according to [80]. For example, the computational complexity for geodesic distance calculation in [4, 70] is $O(mlogm)$, where m is related to a neighborhood area with radius r. Therefore, for n given vertices, the complexity for calculating all geodesic distances is $O(nr^2logr)$. As constants Pi and 2 will be removed from the big O notation. The Euclidean distance, according to [7], is easier to calculate but is sensitive to deformations. When multirings are used, for example like in [6], the geodesic distance between two points on a mesh is approximated properly. They are computationally efficient. We have found that the methods described in [3], [4] and [5]

are exerting more influence on other research works because of their effective results and deformations handling.

2.3.2 Curve-based methods

These methods use a set of curves from facial surfaces as features. The latter include rich geometrical information that captures shape information from different facial regions to represent a 3D face. Compared to keypoint-based methods, they are less sparse and more robust against facial expressions. In addition, the weight of the reference point (often the nose tip) is higher than other points, as it contains descriptive shape informataion. Curve-based methods can be grouped into two categories : contour- and profile-based [31].

Contour-based

Contours are closed curves with different lengths and without intersections. They are defined as level curves classified into iso-depth and iso-geodesic curves. Iso-depth curves, first introduced in [9], are obtained by translating a plane through the facial surface in one direction. These curves are described using the intersections between the facial surface and a plane. For a facial surface S, a set of level curves c_λ is obtained, where each c_λ consists of all points p such that $F(p) = \lambda$, with F being a depth value function for the z component of point p. An extension of this framework is proposed in [8], where level curves of a facial surface distance function, with the origin being the nose tip, are described as iso-geodesic curves. An iso-geodesic curve c_λ consists of the set of all points, whose geodesic distance $dist$, from a reference point r, is in the range $[\lambda - \delta, \lambda + \delta]$, for a small positive δ. A Riemannian analysis framework is employed for comparing facial curves. The latter have the advantage of being invariant to rotations or translations (isometric transformation). However, both iso-depth and iso-geodesic curves (illustrated in figure 2.7), are sensitive to large facial expressions, occlusions and missing parts. Iso-geodesic stripes have also been applied by Berretti et al. [82]. To extract stripes, the normalized geodesic distance $\bar{\gamma}$ is com-

Table 2.2 – Keypoint-based methods

Category/Reference,Year	Database	Matching	Limitation	Advantage (Robustness)
SIFT-like				
Mian et al. 2008[3]	FRGCv2	Graph	Occlusion, partial scans	Expression
Mayo and Zhang 2009[62]	GavabDB	Weighted matching	Complexity, expression	Frontal neutral
Huang et al. 2010[64]	FRGCv2	Hybrid	Noise	Registration-free (frontal), partial occlusion
Berretti et al. 2011[61]	FRGCv2	χ^2 dist	Cost, keypoints redundancy	Partial occlusion
Huang et al. 2012[81]	FRGCv2, Bosphorus, Gavab DB	Hybrid	Large pose (manual landmarks)	Registration-free (frontal)
Inan and Halici 2012[63]	FRGCv2	Cosine dist	Noise	Neutral expression
Soltanpour and Wu 2016[66]	FRGCv2, Bosphorus	Histogram matching	Pose	Expression
Lei et al. 2016[67]	Bosphorus, GavabDB, UMB-DB, SHREC08, BU-3DFE, FRGCv2	Two-Phase Weighted Collaborative Representation	Extreme pose, expression	Partial data, time efficient
Mesh-based				
Li et al. 2011[5]	Bosphorus	Cosine dist	Occlusion, pose, missing parts	Expression
Smeets et al. 2013[4]	FRGCv2, Bosphorus, SHREC11	Angles comparison	Noise	Expression, partial data
Berretti et al. 2014[6]	FRGCv2, BU-3DFE, Bosphorus	Bhattacharyya dist	Noise,verification accuracy	Expression, occlusion, missing parts
Li et al. 2015[70]	Bosphorus, FRGCv2	Fine-Grained Matcher	Cost	Expression, occlusion
Elaiwat et al. 2015[7]	FRGCv2, BU-3DFE, Bosphorus	Cosine dist	Occlusion, missing data	Illumination, expressions
Landmarks				
Koudelka et al. 2005[75]	FRGCv1	Hausdorff dist	Pose	Expression, missing data
Lu et al. 2006[71]	MSU, USF [71]	Hybrid iterative closest point (ICP)	Expression	Pose, light
Colombo et al. 2006[73]	150 3D faces	PCA-based	Noise, artifacts	Pose, expression, light
Gupta et al. 2010[76]	Texas 3DFRD	Euclidean dist	Large expression, pose	Fiducial point localization error
Creusot et al. 2011[74]	FRGCv2	-	Complexity	Expression
Creusot et al. 2013[21]	FRGCv2, Bosphorus	-	Complexity	Expression

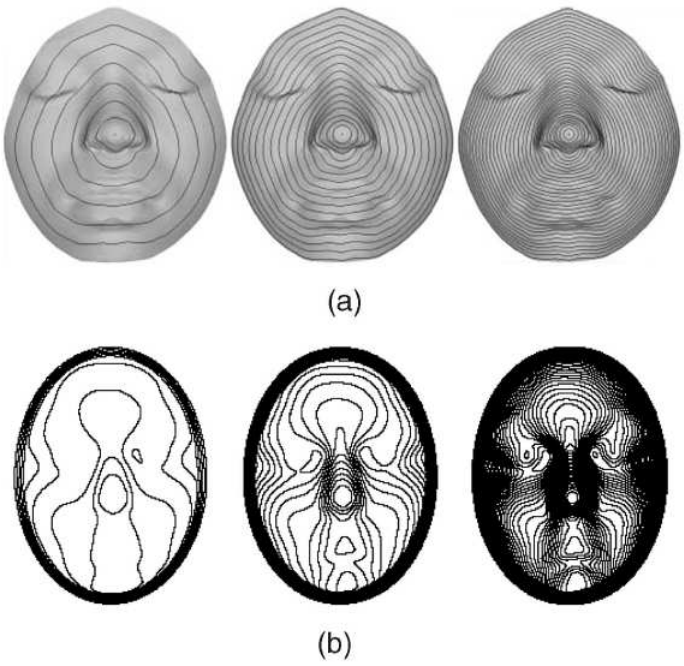

Figure 2.7 – Level curves of a) geodesic function [8], and b) depth function [9] for several levels

puted between each face point and the nose tip, and quantized into N intervals $c_1, ..., c_N$. This way, the i^{th} stripe consists of all points whose distances $\bar{\gamma}$ are in the interval c_i. The stripes are described by a 3D Weighted Walkthroughs (3DWWs) descriptor and used as nodes in a graph-based matching scheme. Level curves have been also employed in [83], [84], [85], [86], and [87]. The main limitation of most of these approaches, apart from occlusion, is their lack of robustness to extremely large facial expressions.

Profile-based

Profiles are open curves, with starting and end points. Typically, the starting and end points are in the middle and on the edge of the face, respectively [31]. Radial curves have been introduced by Drira et al. [88] and extended in [10]. These curves are more efficient than level curves [9], [8], as they cover different face regions that are related to different facial expressions. At least some parts of radial curves are available to handle occlusions

and missing parts. Each curve originates at the nose tip and has an angle α, relative to a reference curve, which is the vertical curve after rotating the face to the upright position. The intersection of the plane p_α and facial surface S yields the radial curve β_α, as shown in figure 2.8. Radial curves are also used together with level sets in [89] to approximate the facial surface. The well-known machine learning algorithm, AdaBoost, is used to select the most efficient features. The machine learning based feature selection method provides a very compact signature of a 3D face and a fast classification approach for face recognition. Using all curves, the computational time for recognition is 2.64s. However, with selected curves the time is reduced to 0.68s, showing that the selection method enhances the system computational performance. Facial curves are widely used as profile-based methods to handle facial expression. Angular radial signatures (ARS) [11] are defined as a set of curves at an interval of θ radians ($\theta \in [0, \pi]$), emanating from the nose tip. A binary mask is defined on the $xy-$plane to project ARSs along different directions. Each resulting path consists of 20 points, with 3mm distance between any two adjacent points. ARS feature value of these points is computed from the depth value of each point, using bicubic interpolation at the x and y coordinates. The ARS extracts significantly a set of discriminative 1D feature vectors from the complex 3D facial surface that achieves computationally efficiency in recognition task. On an INTEL Core 2 Quad-CPU and 8 GB RAM, face identification only requires 6.07s. In particular, the features extracted from semi-rigid regions are robust under facial expressions. Figure 2.9 shows a binary mask, used to extract ARSs, and 17 ARSs on a face.

Another facial curve, obtained by connecting SIFT keypoints, is introduced by Berretti et al. [90] to handle missing data. Because SIFT descriptors are not discriminant enough to recognize an identity, facial curves from pairs of keypoints are defined to create effective features. A graph of facial curves is constructed between matched keypoints. The performance of the method in terms of accuracy can be improved at the curve matching level, when a robust solution is used. In [91] facial curves, the intersection of a plane P and the facial surface are employed to make a rejection classifier. An adaptive region extraction is used for matching two 3D faces. The vertical facial curve in the nose tip is called central

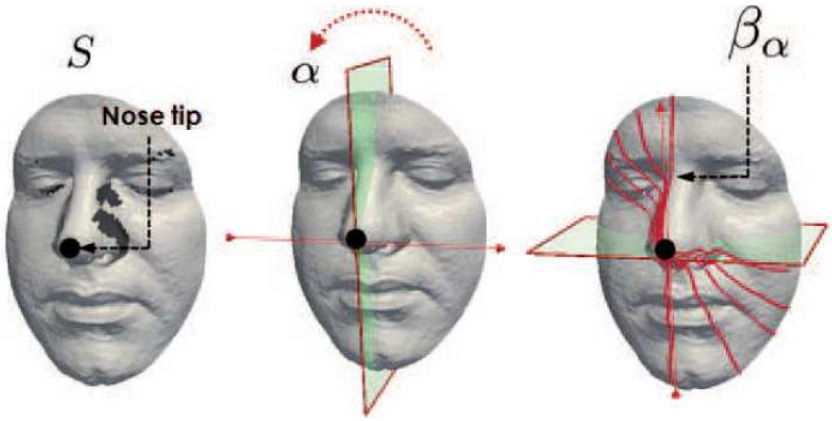

Figure 2.8 – Nose tip, the reference curvature, radial curves [10]

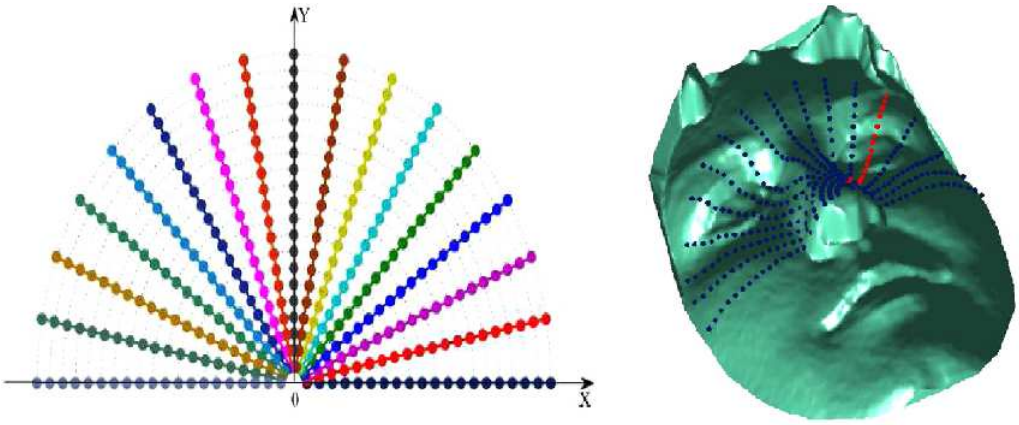

Figure 2.9 – The binary mask and 17 ARSs [11]

profile. Although the partial central profile is less descriptive than the entire one, it is also less sensitive to facial expression and occlusions, less complex, and hence it is used to make a rejector. The similarity between a partial central profile and its corresponding profile from another face is calculated based on the average distance between the two curves, using the iterative closest point (ICP) algorithm [92]. Generally, curves are less discriminative than regions. However, they are faster and require less space for storage. Using an Intel Core Duo 2.34GHz machine with 1GB of memory, the verification process takes less than 9s and recognition process requires 195 s with rejection (608 s without rejection), which shows

that the rejection-based method is faster. Vertical central profile has been also used in [93], where it is defined as the intersection between the symmetry plane, the facial surface and mean curvature. Authors apply the property of the face bilateral symmetry to develop a fast algorithm that on a 1-GHz Pentium IV PC with 512 MB RAM takes an average time equal to 0.5 s for comparison. Recently, some profile-based methods have been proposed that extend the application of facial curves. A set of Rotation-invariant and Adjustable Integral Kernels (RAIKs) is computed from the surface patch around a 3D point, in [94]. Nasal patches and curves are introduced in [95]. First, nasal landmarks are detected, and then using pairs of landmarks, a set of planes is created. The intersection of these planes with the nasal surface yields nasal region curves. These curves are applied to make the feature descriptor. The feature vector is obtained by concatenating histograms of x, y, and z components of the normal vectors of the Gabor wavelet filtered surface. A genetic algorithm (GA) is used to select the more robust features against facial expressions. This method has shown high class separability compared to previous methods.

Summary

Table 2.3 summarizes our surveyed curve-based methods that we divided into contour-based and profile-based categories. In most curve-based methods, the nose tip is used as a reference point or the origin of the system. Since the nose region is rigid, robust under facial expression, and contains more distinctive shape features than other regions, curve-based methods are robust under facial expression. However, hair covering the face, large pose changing, and missing data affect the correct detection of the nose tip. Consequently, face alignment and facial curve extraction are calculated using an incorrect origin, which affects the recognition performance of these methods. In particular, we have found that iso-depth curves [9], iso-geodesic curves [82], and the radial curves [10] are more effective and have greater influence on other researchers.

Table 2.3 – Curve-based methods

Category/Reference,Year	Database	Matching	Limitation	Advantage (Robustness)
Contours				
Samir et al. 2006[9]	FSU, ND [26]	Euclidean,geometric dist	Occlusion, missing parts	Expression
Mpiperis et al. 2007[84]	About 800 images (20 persons)	Euclidean dist	Expression	Translation, planar rotation
Feng et al. 2007[87]	FRGCv2	Cosine dist	Occlusion, missing data	Pose, expression
Mpiperis et al. 2007[86]	BU-3DFE, DB (70 people)	PCA-based	Noise	Large expression
Li et al. 2008[85]	CASIA	Mahalanobis cosine dist	Open mouths	Expression
Jahanbin et al. 2008[83]	1196 images (119 subjects)	Euclidean dist, Support vector machine (SVM)	Occlusion, missing part	Expression
Samir et al. 2009[8]	Laser-scanned	Riemannian framework	Occlusion, missing part	Expression
Berretti et al. 2010[82]	FRGCv2, SHREC08	Graph	Large expression (open mouths)	Identification (large DB)
Profiles				
Zhang et al. 2006[93]	382 faces (166 subjects)	Mean dist	Extreme expression	Fast
Li and Da 2012[91]	FRGCv2	Region-based (ICP)	Large expression, hair occlusion	Expression, time efficient
Ballihi et al. 2012[89]	FRGCv2	AdaBoost	Occlusions	Efficient (data storage, transmission cost), expression
Drira et al. 2013[10]	FRGCv2, GavabDB, Bosphorus	Riemannian framework	Extreme expression, complexity	Pose, missing data
Berretti et al. 2013[90]	FRGCv2, GavabDB, UND	Sparse	Large pose, expression	Missing parts
Lei et al. 2014[11]	FRGCv2, SHREC08	SVM	Occlusion, missing parts	Efficient, expression
Emambakhsh and Evans 2017[95]	FRGC, Bosphorus, BU-3DFE	Mahalanobis, cosine dist	Occlusion	Expression
Al-Osaimi 2016[94]	FRGC, 3D-TEC, Bosphorus	Euclidean dist	Occlusion	Expression

2.3.3 Local surface-based methods

Most local surface-based methods extract local geometric information, from several patches of the facial surface or from some regions of the surface, that are invariant under facial expression variations. These methods can be divided into LBP-based, geometric feature-based and others.

LBP-based

Inspired by the efficient Local Binary Pattern (LBP) for 2D face recognition, LBP-based methods, as surface descriptors, have been developed for expression-robust 3D face recognition. LBP is a local shape descriptor that was initially introduced by Ojala et al. [42] for 2D images. LBP was first employed by Li et al. [96] on intensity image and surface in a fusion scheme for 3D face recognition. Later, 3DLBP [97], in combination with global matching was proposed. A multiscale extended LBP with a SIFT-based strategy is described in [81]. LBP representation is also applied in [98] where, the face division pattern is used to extract depth and normal information encoded by LBP. Local normal pattern (LNP), proposed in [15], encodes facial normal component in the same way as the LBP operator. LNP is defined by the decimal numbers from the encoding process. The histogram-based statistics of LNP values are used as the facial descriptor. To overcome facial expression variations, the weight of each facial normal image patch is learned and applied in a weighted sparse representation-based classifier. The computational complexity of the method depends on the gallery size and feature dimension. The identification time (gallery size: 466) on a PC with Intel Core 2 CPU and 2.66 GHz has been reported at 3.55 s for this method. The results are only reported for face identification and the descriptor is not evaluated in verification tasks. In addition, basic ICP algorithm has been utilized for face registration, showing that applying an efficient registration method can improve the recognition results. Werghi et al. [99] proposed the Mesh-LBP method, where they applied LBP descriptor on mesh. The method was extended to face recognition in [12]. For each

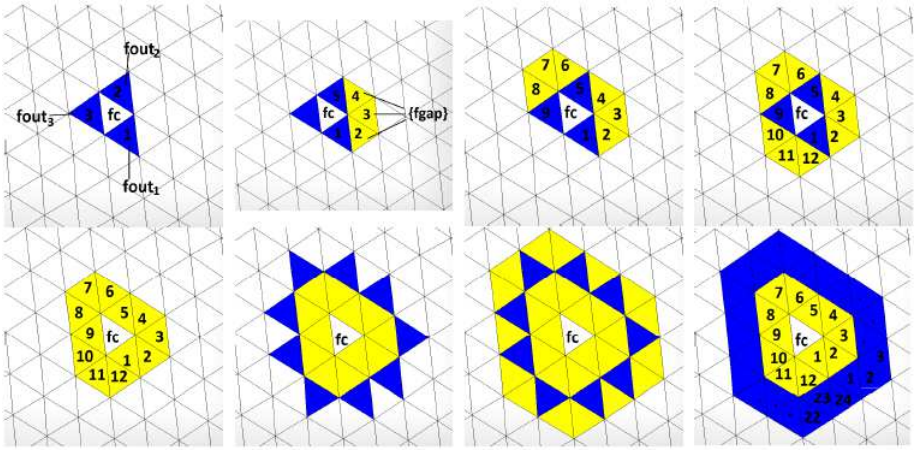

Figure 2.10 – Ordered ring construction, three *Fout* facets adjacent to the f_c and sequences of $Fgap$ facets [12]

central facet f_c on the mesh, *Fout* and $Fgap$ are considered edge facets of f_c. Starting with three *Fout* facets around f_c, the $Fgap$ facets between each pair of *Fout* facets are extracted and the outcome of this procedure is a ring of ordered facets around f_c (see figure 2.10). The mesh-LBP is computed for facet f_c as $\sum_{k=0}^{m-1} s(h(f_k^r) - h(f_c)).\alpha(k)$, with s(x)=1 if x $\geq$ 0 and s(x)=0 if x < 0, where r and m are the ring number and the number of facets on the ring, respectively. The function $h(f)$ is a scalar function that contains either a geometric or a photometric information, such as curvature and color or gray level, respectively. For $\alpha(k)$, two variants are considered, i.e., $\alpha_1(k) = 1$ and $\alpha_2(k) = 2^k$. The curvature maps including curvedness, Gaussian curvature, shape index, and the gray level are used for $h(f)$ in two different shape and texture modalities in a fusion scheme. A constructed histogram over a given neighborhood is considered as a descriptor in the matching step.

Geometric feature-based

Some methods are developed based on geometric features. Xu et al. [100] proposed a 3D face recognition method using geometric features and shape variation information. First, the 3D point cloud is converted to a mesh, then a geometric feature vector is built, using Z-coordinate, $Z(v_i)$ of each vertex v_i of the mesh. Shape features are extracted on some

regions of the face, including mouth, nose, left and right eyes. Two vectors, including geometric and shape features, are concatenated together to make the feature vector. PCA is then applied to reduce the feature dimension. Li and Zhang [101] proposed a recognition system by using geometrical attributes consisting of angles, geodesic distances, and curvatures. To have a stable feature vector under facial expression, expression-insensitive signatures are constructed using weighted attributes. In [102], an expression-insensitive descriptor (EID) based on the sparse representation of low-level geometric information is proposed, where a pooling and ranking scheme is employed to select higher ranked EIDs. Recently, low-level geometric features have been proposed in [13]. These features measure distances and angles between vertices of the 3D mesh. They are robust under facial pose and expression variations as they are calculated for three different regions of the face, viz., semi-rigid (eye-forehead), rigid (the nose) and non-rigid (mouth) regions. Each region is represented using multiple triangles, with one vertex being the nose tip and two randomly selected vertices, from the surface of the region. Using these triangles, low-level geometric features are computed from the angle between the two segments connecting each of the random points to the nose tip (A), the radius of the circumscribed circle (C), the distance between the two random vertices (D), and the angle between the line connecting the two vertices and the z-axis (N) (see Figure 2.11). Each feature vector is normalized into $[-1,+1]$ and quantized into a histogram with m bins. The feature descriptor is calculated by concatenating the four histograms. A support vector machine (SVM) classifier is used to recognize test faces. An extension of this work has been proposed in [67] as local Keypoint-based Multiple Triangle Statistics (KMTS) (Section 3.1). Covariance matrices of descriptors are proposed by Tabia et al. [103] to capture geometric and special properties of a region with the correlation of these properties. For a 3D shape with a set of patches $\{P_i, i = 1...m\}$ around a representative point p_i, a feature vector f_i of dimension d for each point p_j in the patch P_i is computed using $p_j - p_c$, the distance between p_j and p_i, and the volume of the parallelepiped where p_c is the patch center and is equal to $1/n_i \sum_{k=1}^{n_i} p_k$. The representation is generic and other features can be added. A $d \times d$ covariance matrix

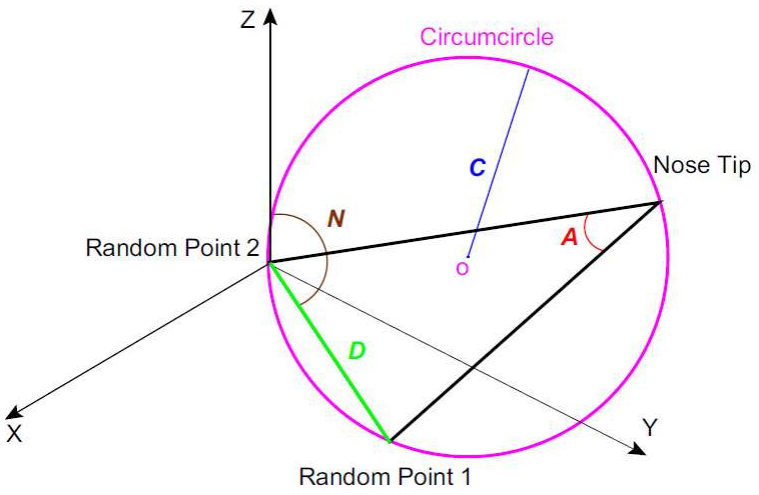

Figure 2.11 – Low-level geometric features [13]

$X_i = 1/n \sum_{j=1}^{n} (f_j - \mu)(f_j - \mu)^T$ is calculated where μ is the mean of the feature vectors f_i. An extension of this work has been presented in [104].

Other methods

Point signature was initially proposed by Chua et al. [105] as a representation for free-form surfaces. To deal with expression variations, only the rigid parts of the face are used in the matching process. Point signature is also used to describe feature points in 3D domain in [106]. Multiple overlapping regions around the nose are extracted using surface curvatures, including mean curvature H and Gaussian curvature K in [107]. An extension of this work is proposed by Flatemier et al.[108]. In [109], the authors introduced tensors, where third order tensors are indexed using 4D hash table. Rank-0 tensor fields are also applied by Al-Osaimi et al. [110], where multiple local tensor fields are computed over a triangular mesh and used as geometrical cues. Most of these methods work based on the surface registration and descriptors that are not suitable for real applications and are computationally expensive. Recently, Ming [111] proposed a regional bounding spherical descriptor that is computationally efficient and handles facial emotions with high recognition rate. This method takes 5.96 s for the whole data processing, which is considered

time efficient. In addition, 2D features can be calculated on 2D maps extracted from 3D meshes to decribe local features such as Gabor filter coefficients in [112]. An extension of this work using wavelet coefficients has been presented in [113]. The authors apply feature scoring to define compact signatures that makes the matching more efficient, especially in large-scale databases. Using an AMD Opteron processor at 2.1 GHz, the algorithm can perform 1,800,000.00 comparisons per second.

Summary

Table 2.4 summarizes the local surface-based methods that we have presented in this section. Some methods such as [15, 12] are inspired by LBP local descriptors with effective performance. Recently, geometric features have been used, for example in [13] and [103], yielding robust descriptors that are capable of handling facial expressions. The methods in [105, 106] use point signatures that are invariant to translations and rotations. Some of the surface-based methods [107, 108] work on some regions of the face that are extracted based on the nose tip location. Hence, they are sensitive to the nose tip detection accuracy. However, these methods are robust under facial expressions. Tensor features, used in [109, 110], combine global and local geometric features and are robust under rigid transformations.

2.4 Discussion

In the past decade, 3D face recognition has significantly grown in terms of databases, features, matching approaches, and even handling degradation conditions. Many 3D face recognition methods rely on local features to overcome deformations.

Tables 2.5 and 2.6 summarize the performances of the surveyed methods, along with their category and performance on different databases under different conditions (expression or neutral (FRGCv2, Bosphorus) pose or frontal (Gavab, Bosphorus)). The criteria used in the literature consist of rank-1 recognition rate (RR1), equal error rate (EER) and verification rate (VR). There are some protocols for experiments on FRGCv2 according to

Table 2.4 – Local surface-based methods

Category/Reference, Year	Database	Matching	Limitation	Advantage (Robustness)
LBP-based				
Li et al. 2005[96]	2305 images (2D+3D)	AdaBoost	Occlusion, missing data	Pose, expression, lighting
Huang et al. 2006[97]	FRGCv2	χ^2 divergence	Pose, occlusion	Expression
Tang et al. 2013[98]	FRGCv2, BJUT-3D [98]	Nearest-neighbor (NN)	Occlusion, missing data	Expression
Li et al. 2014[15]	FRGCv2, Bosphorus, BU-3DFE, 3D-TEC	Sparse-based	Pose, occlusion	Expression, fast
Werghi et al. 2016[12]	BU-3DFE, Bosphorus	Cosine, $\chi^2 dist$	Pose	Expression, missing data
Geometric features				
Xu et al. 2004[100]	3D_RMA [114]	NN	Expression	Pose
Li and Zhang 2007[101]	GavabDB, FRGCv2	NN	Manual feature selection, occlusions	Expression
Li et al. 2009[102]	GavabDB, FRGCv2	Sparse representation	Manual facial markers detection, complexity	Expression
Lei et al. 2013[13]	FRGCv2, BU-3DFE	SVM	Occlusion, missing data	Expression, cost
Tabia et al. 2014[103]	GavabDB	Riemannian framework	Occlusion, missing data	Expression
Hariri et al. 2016[104]	FRGCv2, GavabDB	Geodesic dist	Partial occlusions	Expression, pose
Others				
Chua et al. 2000[105]	6 subjects	Euclidean dist	Misalignment, noise, extreme expression, complexity	Translation, rotation
Mian et al. 2005[109]	UND	Linear correlation	Expression	Pose, no pre-processing
Wang and Chua 2006[106]	80 persons	Structural Hausdorff dist	Hair covering, significant expression	Viewing direction, translation, rotation
Chang et al. 2006[107]	4000 scans (449 subjects)	ICP-based	Find landmarks, hair on the face	Expression
Faltemier et al. 2008[108]	FRGCv2	ICP-based	Non-frontal, incomplete data	Expression
Al-Osaimi et al. 2008[110]	FRGCv2	PCA-based	Hair covering, noise	Mild expression
Ocegueda et al. 2013[113]	FRGCv2, BU-3DFE, Bosphorus	Linear Discriminant Analysis (LDA)-based	Pose, occlusion	Expression
Ming 2015[111]	FRGCv2, CASIA, BU-3DFE	Regional, global regression	Patches detection	Large pose, efficient

Table 2.5 – Performance of local feature methods on FRGCv2 Database: 0.1% FAR VR and RR1 (”n/n: neutral vs. neutral”, ”n/a: neutral vs. all”, ”n/nn: neutral vs. non-neutral”, ”a/a: all vs. all”)

Category/Reference,Year	VR(n/n)	VR(n/a)	VR(n/nn)	VR(a/a)	VR(ROC III)	RR1(n/a)	RR1	Computational performance
Keypoints								
Lei et al. 2016[67]	99.9%	98.3%	96%	-	-	96.3%	n/n:99.6%, n/nn: 92.2%	Identification: 1.82 s(gallery size: 466) with an Intel Core 2 Quad CPU and 16 GB RAM
Soltanpour and Wu 2016[66]	99.9%	99.3%	98.4%	99%	98.65%	96.9%	n/nn: 96%	one match: 0.35 s, Intel Core i7 3.60 GHz CPU with 8 GB RAM
Elaiwat et al. 2015[7]	99.9%	a/n: 99.2%	nn/n: 98%	-	97.8% (FAR=o.1)	a/n: 97.1%	n/n: 99.4%, nn/n: 94.1%	One match: 0.36 s Intel Core i7 3.40 GHz processor with 8 GB RAM
Li et al. 2015[70]	-	-	-	-	-	96.3%	-	-
Berretti et al. 2014[6]	-	-	-	-	86.6%	-	-	-
Smeets et al. 2013[4]	-	-	-	79%	77.2%	-	89.6%	-
Huang et al. 2012[81]	99.6%	98.4%	97.2%	94.2%	95.0%	97.6%	n/n: 99.2%, n/nn: 95.1%	One match: 0.32 s, Intel(R) Core(TM) i5 CPU (2.60 GHz) and 4 GB RAM
Inan and halici 2012[63]	-	-	-	98.35%	98.25%	97.5%	-	Time complexity: O(n), n is the number of processed 3D scans
Berretti et al. 2011[61]	-	-	-	-	-	-	Partial faces: 89.2%	-
Huang et al. 2010[64]	-	-	-	-	-	96.1%	n/n: 99.1%, n/nn: 92.5%	65 matches/s: 2.66 GHz Pentium IV machine with 4 GB RAM
Mian et al. 2008[3]	99.9%	98.6%	96.6%	-	-	96.1%	n/nn: 92.1%	
Koudelka et al. 2005[75]	-	-	-	-	-	FRGCv1: 94%	-	Feature extraction: 2.5 s (average) 3.4 GHz Pentium 4
Curves								
Emambakhsh and Evans 2017[95]	-	-	-	-	93.5%	97.9%	n/nn: 98.5%	-
Al-Osaimi 2016[94]	99.8%	98.69%	nn/n: 98.25%	-	-	a/n: 97.78%	nn/n: 96.49%	Local surface: 45 μs (average) 12-core machine (c/c++)
Lei et al. 2014[11]	-	-	97.8%	-	96.7%	-	-	Identification (n/nn): 6.07 s, INTEL Core 2 Quad CPU and 8 GB RAM
Drira et al. 2013[10]	-	-	-	93.96%	97.14%	97.7%	97%, n/nn: 96.8%	Match: 1.27 s
Berretti et al. 2013[90]	-	-	-	-	-	95.6%	n/n: 97.3%, n/nn: 92.8%	One match: 0.2 s, Centrino Duo 2.2 GHZ with 2 GB memory
Li and Da 2012[91]	-			95.3%	96.0%	97.8%	-	Total cost: 195 s, Intel Core Duo 2.34 GHz 1 GB of memory
Ballihi et al. 2012[89]	-	-	-	-	-	98.02%	-	Identification: 0.68 s (gallery size: 466)
Berretti et al. 2010[82]	97.7%	a/n: 95.5%	nn/n: 91.4%	81.2%	-	-	-	-
Local surfaces								
Hariri et al. 2016[104]	-	-	-	-	-	99.2%	-	-
Ming 2015[111]	-	-	-	-	95.03%	-	-	Identification: 5.96 s
Li et al. 2014[15]	-	-	-	-	-	96.3%	nn: 94.2%	Identification: 3.55 s, Intel Core 2 CPU and 2.66 GHz
Lei et al. 2013[13]	-	-	97.6%	-	-	-	n/nn: 95.6%	-
Tang et al. 2013[98]	-	-	-	-	-	-	94.89%	-
Ocegueda et al. 2013[113]	-	-	-	-	97.5%	96.6%	-	Comparisons/s: 1,800,000.00 AMD Opteron 2.1 GHz
Li et al. 2009[102]	-	-	-	-	-	93%	n: 93.33%, nn: 92.78%	-
Faltemier et al. 2008[108]	-	-	-	93.2%	94.8%	98.4%	n or nn/rest: 97.2% n/nn: 96.3%	Verification < 10 s Intel P4 2.4 GHz
Al-Osaimi et al. 2008[110]	95.37%	-	-	-	-	-	n/n: 93.78%	-
Li and Zhang 2007[101]	-	-	-	-	-	-	RR(180 faces)=99.45%	-
Huang et al. 2006[97]	-	EER=9.4%	-	EER=10%	EER=10.7%	-	-	-
Chang et al. 2006[107]	EER=0.12	-	EER=0.23	-	-	-	n/n: 97.1%, n/nn: 87.1%	-

Table 2.6 – Performance of local feature methods on other databases: 0.1% FAR VR and RR1 (”e: expression”, ”n: neutral”, ”nn: non-neutral”, ”a: all”, ”o: occlusion”, ”p: pose”)

Category/Reference,Year	Bousphorus	GavabDB	BU-3DFE	SHREC08/11	OtherDB
Keypoints					
Lei et al. 2016[67]	RR1(e,o,p)=98.9%, 94.2%, 90.6%	RR1(a)=96.99%	VR(n/a)=94%	08: VR(n/a)=93.91%	UMB-DB: RR1 (n/o)=73.08%
Soltanpour and Wu 2016[66]	VR,RR1(n/e,o)=98.4%,97.2%, VR,RR1(n/a)=95.8%,94.5%	-	-	-	-
Elaiwat et al. 2015[7]	VR=91%	-	VR(a/n)=95.01%	-	-
Li et al. 2015[70]	RR1(n/a)=96.56%	-	-	-	-
Berretti et al. 2014[6]	RR1(n/a)=94.5%	-	RR1=88.2%	-	-
Smeets et al. 2013[4]	RR1(n/a)=93.7%	-	-	11:RR=98.57%	-
Huang et al. 2012[81]	RR1(n/(e,o))=97.0%	RR1(no pose)=95.49%, (pose)=91.39%	-	-	-
Li et al. 2011[5]	RR=94.10%	-	-	-	-
Gupta et al. 2010[76]	-	-	-	-	1149 scans:EER=1.98%, RR1=96.8%
Mayo and Zhang 2009[62]	-	RR(n=95%,nn=90%),	-	-	-
Lu et al. 2006[71]	-	-	-	-	MSU,USF:RR1=90%
Colombo et al. 2006[73]	-	-	-	-	150 faces:RR=82%
Curves					
Emambakhsh and Evans 2017[95]	RR1(n/n,e)=98.96%, 95.35%, RR1(n/n)=	-	RR1(n/nn)=88.9%	-	-
Al-Osaimi 2016[94]	RR1(n/a,e,o)=90.28%,92.41%, 84.78%	-	-	-	3D-TEC: RR1(n/e)=85.98%
Lei et al. 2014[11]	-	-	-	08:VR(n/nn)=88.5%	-
Drira et al. 2013[10]	RR(n/o)=87%	RR=96.99%	-	-	-
Berretti et al. 2013[90]	-	RR1(frontal nn,n+nn)= 96.17%, 97.13%	-	-	UND: RR1(144 scans)= 75%
Berretti et al. 2010[82]	-	-	-	08:VR(a/n,a/a,nn/n)=90.4%, 80.63%, 80.87% RR=99.53%	- - -
Jahanbin et al. 2008[83]	-	-	-	-	1196 images:EER=2.58%
Li et al. 2008[85]	-	-	-	-	CASIA: RR1(Opened mouth +e)=85.3%
Mpiperis et al. 2007[84]	-	-	-	-	about 800 images:RR= 91.36%
Mpiperis et al. 2007 [86]	-	-	RR1=84.4%	-	-
Samir et al. 2006[9]	-	-	-	-	FSU:RR=92%, ND:RR1 =90.4%
Zhang et al. 2006[93]	-	-	-	-	382scans:RR1(n,e)=96.9%, 87.5%
Local surfaces					
Werghi et al. 2016[12]	RR1(occlusion)= 99.38%	-	RR1(e)=93.42%	-	-
Hariri et al. 2016[104]	-	RR=97.81%	-	-	-
Ming 2015[111]	-	-	-	-	CASIA: RR1(p,e)=82%, RR1(p)=92.5%
Li et al. 2014[15]	RR1=95.4%	-	RR1=92.21%	-	3D-TEC: RR1=95.3%
Tabia et al. 2014[103]	-	RR=94.91%	-	-	-
Lei et al. 2013[13]	-	-	VR=98.2%, RR1=97.7%	-	-
Tang et al. 2013[98]	-	-	-	-	BJUT-3D: RR (a)=92.4%
Ocegueda et al. 2013[113]	VR(n/a, a/a)=93.8%, 92.2%	-	VR(n/a, a/a)=96.3%, 94.8%	-	-
Li et al. 2009[102]	-	RR(n, e, a)=96.67%, 93.33%, 94.68%	-	-	-
Li and Zhang 2007[101]	-	RR(frontal)=97%	-	-	-
Mian et al. 2005[109]	-	-	-	-	UND:RR1=86.4%
Xu et al. 2004[100]	-	-	-	-	3D_RMA (120 persons): Classification R=72.4%

[16], and most authors reported the results by following these protocols. Hence, table 2.5 has been assigned to performances on FRGCv2 database, including VR and RR1 for the experiments based on the neutral and non-neutral sets of databases (neutral vs. non-neutral, vs. neutral, vs. all, and all vs. all) and ROC III experiments [16]. Some earlier papers for example, [105, 106], present their algorithms without any quantitative results. Because different experiments in the literature are presented in various situations and for different conditions and databases, it is difficult to perform an overall fair comparison between all these different methods.

All local 3D face approaches surveyed in this chapter are divided into keypoint-based, curve-based, and surface-based. The keypoint-based category is successful in handling occlusions and missing data [61, 64, 67, 4, 6, 70, 71, 73]. The main disadvantage of these methods is their sparseness that makes them sensitive to noisy data and extreme expression changes [62, 63, 64, 67, 4, 6, 71, 73, 76]. Moreover, the high computation cost of some of the SIFT- and curvature-based methods is another drawback for this category [61, 4, 70, 74]. The comparison of keypoint-based methods shows that some of them work effectively in fusion scheme. That is when combining 3D and intensity images in multimodal mode, resulting in high recognition rate [3, 66, 7]. In addition, meshSIFT [4], meshDoG [6], and keypoint detector using PCA [3] extract the distinct descriptors from the patches around keypoints and benefit from the advantages of the local surface-based category.

The second category, curve-based methods, considers contours and profiles. Although curves are less sparse than keypoints, some parts of the face shape can be missing [31]. Hence, most of the methods in this category are not robust against occlusion and missing data [8, 83, 87, 89, 11, 91, 94, 95]. Most of the profile-based methods, for instance [89, 11, 91], are computationally efficient and handle facial expressions. Geodesic representation of the facial surface describes the invariant properties under isometric deformations. Therefore, iso-geodesics in [8, 82, 83, 85, 86] provide expression robust recognition systems. When the nose tip detection is done properly, curve-based methods are more reliable than other methods, under expression variations. The methods, described in [10, 90]

and evaluated on the Gavab database (with scans under pose variations), propose pose-invariant descriptors, but they are still sensitive under large pose variations.

Surface-based methods, our third category, are based on extracted geometrical invariant descriptors. Most of these methods are expression invariant, because they rely on extracted features from regions, that are relatively stable under facial expressions [13, 12, 107, 108], or they rely on sparse representation to learn weights of expression-insensitive patches and high-ranked features selection [15, 102], or use covariance matrices with geodesic metrics [103, 104]. However, methods in this category are sensitive to occlusions and missing data [13, 15, 101, 103, 104, 106, 108, 110].

Among the more recent works, [13] uses low-level geometric features and is computationally efficient, since it involves only basic computations, such as angles and distances. The method proposed by Li et al. [15] is inspired by the computationally efficient LBP descriptor on surface normal component, and hence provides acceptable cost. Furthermore, 3D face verification, using the method in [111], drastically reduces the computational cost because of its efficient pre-processing and alignment steps, that are done with a simple implementation. The methods that use ICP to perform matching, like [71, 107, 108] from tables 2.2 and 2.4, have a good recognition performance, but are not computationally efficient. However, [91] is an exception in this category, as it uses ICP-based matching but still provides an efficient classification, because of its rejection classifier that quickly eliminates dissimilar samples.

Some recent works select the most discriminative features to improve the recognition performance. They use feature selection methods such as AdaBoost, a machine learning technique, [89], a genetic algorithm-based selector [95], sparse representation learning-based method [15], and learning technique like PCA [10].

In particular, this survey suggests that no existing algorithms can handle all existing challenges, including facial expressions, pose variations, occlusions, missing data, hair covering part of the face and background clutter. Incomplete facial data and artifacts are still major issues in practical application of local surface-based methods. Deep learning

might be used to boost the various local feature extraction methods, further improving recognition performance. The latter will be also improved by applying fusion methods, that use 3D data and texture images. Furthermore, applying powerful feature selection methods to find a subset of the most discriminant features is another way to improve the performance of face recognition

2.5 Conclusion

3D face recognition is a vibrant and popular research area in the computer vision and image processing field. Face recognition falls in the category of non-rigid object recognition, where handling deformations effectively still needs improvement. Compared to intensity images, 3D images are more robust against viewpoint and illumination variations, as they contain the local geometry of the face. The challenges in this field such as computational cost reduction and 3D data acquisition techniques enhancement require more work in the future. This survey reviewed recent advances in 3D face recognition, focusing mainly on methods that are based on local features. A taxonomy of the 3D local feature-based methods has been presented in this chapter, together with their advantages and limitations. Properties, including descriptiveness, robustness, compactness and computation efficiency, are important criteria when comparing the effectiveness and strength of each descriptor. Future work could include a comparative study of different local feature extractors for 3D face recognition. We hope this survey will further motivate the researchers in this area to dedicate more consideration and attention to the use of 3D local features for face recognition.

Chapter 3

Multimodal Face Recognition using Local Descriptors

In this chapter, we propose a local descriptor based multimodal approach to improve face recognition performance. Pre-processing is done to smooth, re-sample, and register data. The re-sampled three-dimensional (3D) face data are applied to extract novel descriptors including pyramidal shape index, pyramidal curvedness, pyramidal mean, and Gaussian curvatures. Proposed pyramidal shape maps are extracted at each level of the Gaussian pyramid on each point of the 3D data to have 2D matrices as representatives of 3D geometry information. A local descriptor structural context histogram, which represents the structure of the image using scale invariant feature transform, is calculated on pyramidal shape map descriptors and texture image to find matched keypoints in 3D and 2D modality, respectively. Score-level fusion by means of sum rule is employed to get a final matching score. Experimental results on the Face Recognition Grand Challenge (FRGCv2) database illustrate verification rates of 99% and 98.65% at 0.1% false acceptance rate for all versus all and ROC III experiments, respectively. On Bosphorus database, verification rate of 95.8% for neutral versus all experiment has been achieved.

3.1 Introduction

Hybrid matching is related to the fusion of holistic and feature-based methods or integration of two modalities, 2D facial images and 3D facial surfaces. Multimodal methods are supported by many researches to enhance face recognition. In [115], multimodal method based on the biological vision-based facial description, perceived facial images, with SIFT-

based matching is presented. Weights learning for score-level fusion is done by a genetic algorithm. Keypoint detection based on the curvelet transform on textured 3D face surfaces is employed in [7]. Feature-based multimodal method is presented in [3] based on keypoint detection and fitting a surface to the neighbourhood of a keypoint using a PCA: principal component analysis, subspace of features and SIFT matching in 2D domain. Score and feature level fusion are employed to combine 2D and 3D results. A multimodal method employing scale space extreme on shape index (SI) and texture images is proposed in [116]. Al-Osaimi et al. [117] described a method for combining texture and shape data in a data level fusion approach. Optimisation of fusion function is done to enhance learning capability. A hybrid feature-based and holistic matching with a 3D spherical face representation and SIFT descriptor is used in [22]. In addition, fusion-based approaches are employed in many works to enhance the recognition accuracy. Huang et al. [81] presented a multiscale extended local binary pattern with SIFT-based matching using hybrid matching scheme. Fusion of low-level geometrics features, region-based histogram descriptor, extracted from eye and nose regions with support vector machine as a classifier is proposed in [13]. Score-level and feature-level schemes have been tested and compared. The studies show that the combination of texture and depth information increases face recognition accuracy by making the algorithms robust against degradation conditions [29]. Accordingly, we propose a novel approach to improve recognition accuracy through the conjunction of 2D and 3D face data.

The organization of this chapter is as follows: Section 2 reviews the proposed method; Section 3 elaborates the pre-processing, region of interest (ROI) extraction, and noise removal; Section 4 presents structural context; Section 5 describes pyramidal shape maps of 3D scans and feature extraction on the extracted 2D matrices; Section 6 discusses the matching approach; Section 7 presents experimental results; and Section 8 draws the conclusion.

3.2 Overview of the proposed method

In this chapter, a multimodal local feature-based face recognition algorithm is presented that performs better than holistic algorithms. The block diagram of the proposed method is illustrated in figure 3.1. Nose tip detection is done to extract ROI of face image. Noise and spikes are removed during pre-processing step. Pyramidal shape maps are proposed for 3D recognition extracted using estimation curvature on triangular mesh. Histogram of structural context [118] is calculated on SIFT [41] keypoints on texture image and pyramidal shape maps that works more efficient than SIFT descriptor to find matched keypoints. Geometric attribute based descriptors, local surface patches such as SI, curvature, and so on, are applied as 3D local features in the literatures and have proven to be successful [60]. The 2D features, SIFT descriptors, on 3D SI map have been applied for range image recognition in [40] and for 3D face recognition in [63]. In addition, Huang et al. [64] reported local feature hybrid matching using SIFT descriptors on SI and local binary patterns with successful results in face recognition application. However, these methods are sensitive to the noise. To handle the sensitivity of the existing methods against noise, we propose to extract shape maps on three different scales of the Gaussian pyramid. The proposed pyramidal shape map improves shape information in 3D domain and causes to highly repeatable and robust keypoint identification. In 2D modality, to present more discriminant descriptor than SIFT, we apply histogram of structural context [118] that is invariant to intra-class variation, illumination, noise, rotation, and view point change. Furthermore, we apply the descriptor on pyramidal shape maps obtained from range images to compute similarity between two faces. The numbers of matched keypoints are considered as the matching score in each 2D and 3D face recognition phase and combined by score-level fusion as final score. We test our proposed algorithm on two very famous and challengeable 3D face databases, the Face Recognition Grand Challenge (FRGCv2) [16] and Bosphorus [2]. Preliminary results of this work presented in [65] report only the results for pyramidal SI descriptor on the FRGCv2 database for ROC III and all versus all experiments according to FRGC program

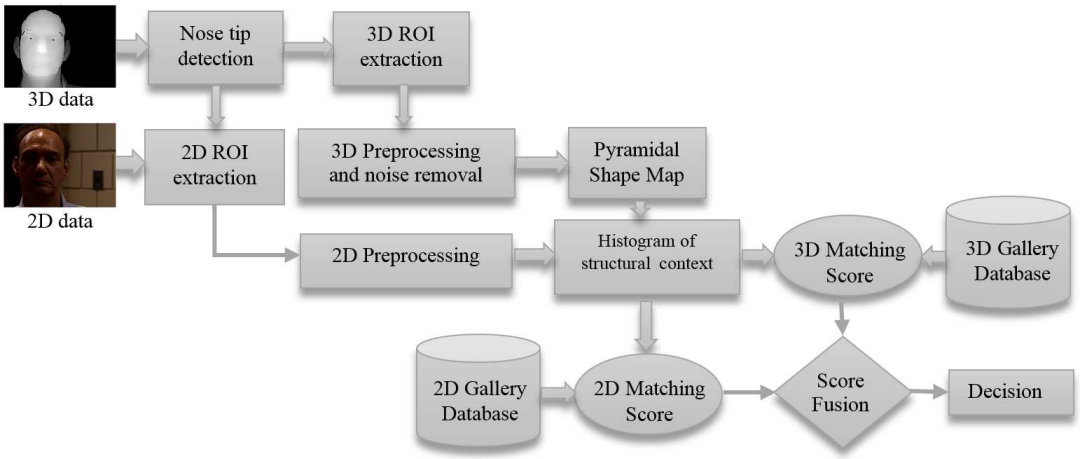

Figure 3.1 – Block diagram of the proposed method

[16]. Different shape descriptors including pyramidal curvedness, mean, and Gaussian curvatures have been proposed in this work and extensive experiments are carried out for each of pyramidal curvedness, mean, and Gaussian curvatures on two databases under different conditions including facial expression, pose variations, and partial occlusion.

3.3 Pre-processing

The first step in biometrics recognition systems is pre-processing of the data, which is an essential and unavoidable task. The proposed algorithm carries out the following tasks in the pre-processing stage.

i. Nose tip detection and cropping

The approach based on [25] is utilized to extract 3D ROI and its corresponding 2D scan using nose tip detection. The position of smallest depth (maximum z-value) for each row is detected, and by computing the number of positions for each column a histogram is created. The peak of the histogram shows the column including the nose tip, and the position with the maximum z-value in this column is detected as the nose tip. The ROI of the corresponding 2D image is extracted by considering the corresponding pixel of the

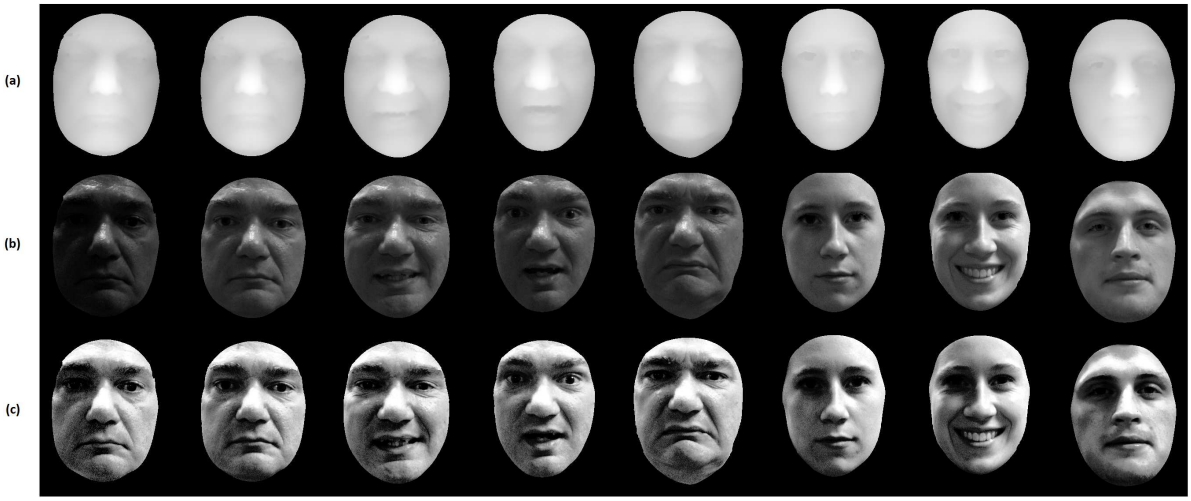

Figure 3.2 – a) Three-dimensional ROI extraction and its corresponding 2D ROI b) before and c) after pre-processing

face oval. To crop the facial surface from the 3D data, a sphere with a radius of 100 mm and with the centre at nose tip is considered, and in recognition only these points are used. We resample and interpolate 3D data at a uniform square grid in the XY-plane at a 1 mm resolution and 400×320 grid size.

ii. Noise removal

Spikes created by sensors are eliminated by means of thresholding and interpolating during the resampling phase. We employ 2D Wiener filtering on the z-component of point clouds as in [38]. To enhance the extracted 2D face, a histogram equalisation algorithm is applied. 3D ROI extraction and its corresponding 2D ROI after and before pre-processing are displayed in figure 3.2.

iii. Orientation correction

Orientation correction is carried out based on the approach presented in [119]. The symmetry axis of the SI map is used to remove rotation in the plane by positioning the detected nose tip at the origin.

3.4 Structural context

A novel approach for face recognition is proposed using structural context [118]. The descriptor is similar to shape context [120] by means of capturing the relationship between the current and remaining points that can represent the structure of the image. In the proposed system, histogram of structural context is applied on 2D texture image and 3D maps to find similar subjects. The first step in calculating the structural context is to extract SIFT [41] keypoints. The Difference of Gaussian (DoG) function, an approximation of the normalized Laplacian, is convolved with the image and sampled. Keypoints are detected as local minimal or maximal of the DoG function. Unreliable points are removed by thresholding. To compute a histogram of structural context around each keypoint, the approach presented in [118] is employed. To make a rotation invariant descriptor, structural orientation is assigned. An orientation histogram is constructed with 36 bins to cover orientations of the interest point for each 10 degree. According to the orientation of interest points, the sum of the scale value of the interest points that fall into each bin is the value of the bin. The peaks in the orientation histogram represent structural orientation. Then, the coordinates of the descriptor and interest point orientation are rotated relative to the structural orientation. Outlier keypoints are eliminated by calculating mean distance of the keypoint to other keypoints and comparing it with the mean distance between all keypoint pairs. If the former one is 30% larger than the later one, the keypoint is outlier and it is eliminated. After elimination, the structural context is constructed by a 5×12 histogram as shown in figure 3.3, the radius of the log-polar to compute the histogram is r/16, r/8, r/4, r/2, r, in which r is 2 after the scale normalization. Each bin of a log-polar histogram is the sum of all scale values of the points in the bin. The structural context descriptor is calculated according to the following equation

$$h_i(k) = \frac{s(p_i)}{max_s} \sum_{p_j \in bin_i(k)} s(p_j) \tag{3.1}$$

In the above equation, $s(p_i)$ and max_s are the scale value of point p_i and the largest

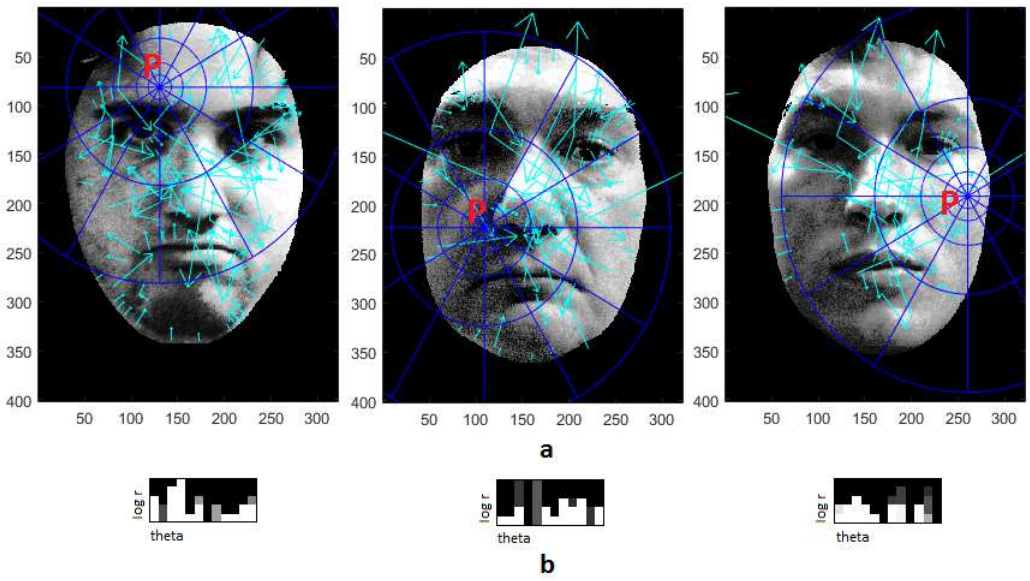

Figure 3.3 – a) Sample keypoint P, b) construction of histogram of structural context

scale of the interest points, respectively. The first part of equation 3.1, $s(p_i)/max_s$ is used for normalization. Since structural context orientation is computed based on the orientation of all the interest points and orientation correction is done in pre-processing, the descriptor is rotation invariant. Since interest points (DoG) are robust in illumination changes, the structural context is illumination invariant.

3.5 Pyramidal shape map

In this section, a new descriptor is proposed and calculated on the 3D shape maps of the face. The normal vector is the unit vector that emerges from the point on 3D space and is perpendicular to the surface. The plane that contains the normal vector is called a normal plane. The 3D curve is the intersection of the normal plane and the surface. There are different curvatures related to the 3D curves in various normal planes. The minimum and maximum values of these curvatures are principle curvatures denoted as k_{min} and k_{max}, respectively. Shape maps including SI, curvedness, mean, and Gaussian curvatures are cal-

culated by means of principle curvatures. There are three different approaches to calculate these curvatures which consist of local fitting, discrete estimation of curvature, and estimation of curvature tensor. In this chapter, we apply the local cubic order fitting method to extract the shape maps from 3D face data, which has better performance on facial expression [5]. In this method, local 3D coordinates frame for each vertex p of a triangular mesh with the origin at p and the normal vector of the vertex $n_p = (n_x, n_y, n_z)^T$ as z-axis is determined. Two orthogonal axes, x and y, are chosen in the tangent plane perpendicular to the normal vector. The average of the normal vectors of the faces adjacent to the vertex is considered as vertex normal. A cubic polynomial function and its normal are represented in equations 3.2 and 3.3. The least-square fitting method is used to solve the fitting equations 3.2 and 3.3. To calculate maximum and minimum curvature, eigenvalues of the Weingarten matrix equation 3.4 are calculated and estimated as principle curvatures.

$$z(x,y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2 + Dx^3 + Ex^2y + Fxy^2 + Gy^3 \tag{3.2}$$

$$(z_x z_{y'} - 1) = (Ax + By + 3Dx^2 + 2Exy + Fy^2 + Bx + Cy + Ex^2 + 2Fxy + 3Gy^2 - 1) \tag{3.3}$$

$$W = \begin{bmatrix} \frac{\partial^2 z(x,y)}{\partial x^2} & \frac{\partial^2 z(x,y)}{\partial x \partial y} \\ \frac{\partial^2 z(x,y)}{\partial x \partial y} & \frac{\partial^2 z(x,y)}{\partial y^2} \end{bmatrix} \tag{3.4}$$

SI [121] which is used to describe local shape topography by calculating curvature on triangular mesh is defined as 3.5 at each point p from the 3D surface computed using maximum curvature k_{max} and minimum curvature k_{min}, for which ($k_{max} > k_{min}$). The value of the SI is between 0 and 1.

$$SI = \frac{1}{2} - \frac{1}{\pi} arcttan \frac{(k_{max} + k_{min})}{(k_{max} - k_{min})} \tag{3.5}$$

Curvedness (C) [121] also calculates local shape topography using the following equation:

$$C = \sqrt{\frac{{k_{min}}^2 + {k_{max}}^2}{2}} \tag{3.6}$$

The mean of the principle curvatures is called mean curvature (H) [121] and is equal to

$$H = \frac{k_{min} + k_{max}}{2} \tag{3.7}$$

The product of principle curvatures is called Gaussian curvature (K) [121] and is calculated as

$$K = k_{min} \times k_{max} \tag{3.8}$$

Pyramidal shape map is the proposed descriptor extracted from 3D data, which creates images at several levels of scales. We use the Gaussian pyramid to decompose 3D data into information at multiple scales, to extract shape map features, and to attenuate noise. At each level, the size or scale of the image is equal to half of the scale of the previous level. The representation consists of two basic operations: smoothing, which works using a sequence of smoothing filters, each of which has twice the radius of the previous one, and down-sampling that reduces image size by half after each smoothing. In this work, the shape map operators including SI, C, H, and K are computed at three levels of the pyramid to capture the whole face and face parts. The shape features at the coarse level of the Gaussian pyramid capture high strength shape of the face and at the fine level capture short and low strength shape information or small details of the triangular mesh from 3D face data. To extract significant information from 3D data, the pyramidal shape map is calculated as follows:

- The Gaussian pyramid of the ROI of the 3D face data, x, y, and z, is created at three levels of scales: 1, 0.5, and 0.25 in the sizes of 400×320, 200×160, and 100×80.

Figure 3.4 – a) 3D shape maps and b) three-level Gaussian pyramid of the SI, C, H, and K descriptors

Thus, the 400×320 image size is considered as the original size of the ROI image and is the fine level of the pyramid. The shape operators (3.5, 3.6, 3.7, and 3.8) are used at each level to extract the local shape map.

- The shape map images are interpolated into the original size. Therefore, three shape map descriptors are extracted at the same size.

- All three local shape descriptors are added together. The resulting matrix is a pyramidal shape map feature. The values of the pyramidal shape map are related carefully to the importance (strengths and shapes) of face structure.

An example of 3D shape maps, including SI, curvedness (C), mean, and Gaussian curvatures (H and K) are illustrated in figure 3.4a. Three levels of the Gaussian pyramid of shape map descriptor operators are represented in figure 3.4b. As the first two of the three levels figures in part b illustrate, SI and curvedness contain more significant features relative to the second two figures, which are mean and Gaussian curvatures.

3.5.1 Feature extraction on pyramidal shape maps

Similarity of two faces are measured using structural context descriptor [118] and finding accurate matched SIFT keypoints. The structural context is calculated on our novel proposed 3D descriptor, pyramidal shape maps. We assume that by occurring a change in facial expression, some small local areas such as nose and the areas around that change slightly and keep invariant. In this way, these regions are utilized in 3D face recognition to handle facial expression variations. Accordingly, our hybrid method, structural context on pyramidal shape maps, is robust to expression variation. The pyramidal shape maps provide the larger numbers of SIFT keypoints that enhance matching performance compared with texture face images and shape maps. It guarantees keypoints repeatability and provides more sufficient distinct features. Keypoints appear at nearly the same location in two different samples of the same person. According to our experiments, descriptors extracted from the pyramidal SI have a larger number relative to pyramidal curvedness, pyramidal mean, and Gaussian curvatures. In our experiments, the average number of the keypoints obtained from 2D texture face images, before and after histogram equalisation, is 40 and 125. From the SI map and pyramidal SI map, 925 and 1430 keypoints are obtained. Figure 3.5a shows keypoints extracted from texture image, the SI map, and the pyramidal SI map. A histogram of structural context descriptor is calculated on each SIFT keypoint according to [118] and applied to the matching process. A histogram of structural context descriptor for a sample keypoint P is displayed in figure 3.5b for pyramidal SI and pyramidal curvedness maps. As this figure depicts, the number of keypoints increases by means of pyramidal shape maps significantly, and our proposed descriptor improves recognition performance.

3.6 Matching approach

This part consists of two phases: histogram of structural context matching on texture images and pyramidal shape maps. The number of matched keypoints is considered to be the

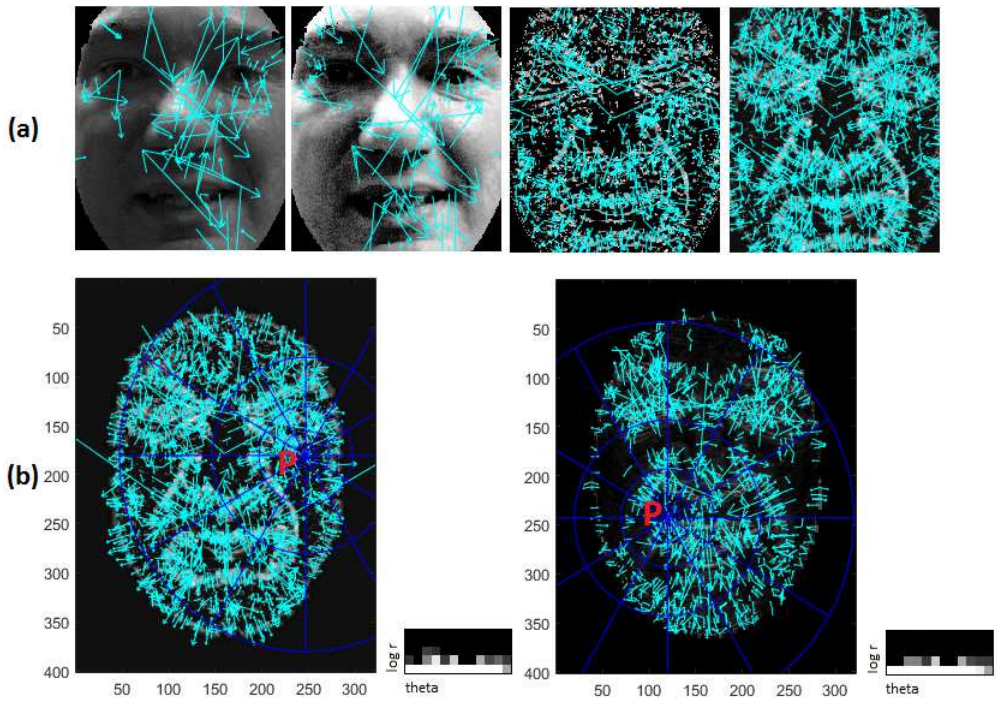

Figure 3.5 – a) Keypoints extracted from the texture images, SI, and pyramidal SI maps and b) histogram of structural context from the pyramidal shape index and curvedness

matching score at each phase. The final score is calculated by means of score-level fusion.

3.6.1 Texture image

The histogram of structural context for each 2D face image from gallery set is extracted and saved as template in the 2D database. For each 2D face image from the probe, a histogram of structural context is extracted and compared with all templates saved in the database to find the number of matched keypoints as the 2D matching score S'_{sc} using the following equation:

$$S'_{sc_{ij}} = \frac{1}{2}\sum_{k=1}^{K}\frac{\left[h_{2D_i}(k) - h_{2D_j}(k)\right]^2}{h_{2D_i}(k) + h_{2D_j}(k)} \tag{3.9}$$

where h_{2D_i} and h_{2D_j} are structural context histograms extracted from 2D images of the gallery and probe sets and K is the number of histogram bins. Matched keypoints using a

comparison of structural context for two texture images of the same subject and different subjects are illustrated in figure 3.6a.

3.6.2 Pyramidal shape map

Histogram of the structural context on the pyramidal shape maps is employed to calculate the 3D matching score S'_{sm} (number of matched keypoints). To detect similarities between two 3D face samples, differences between probe and gallery descriptors are calculated using 3.10 as the 3D matching score S'_{sm}

$$S'_{sm_{ij}} = \frac{1}{2}\sum_{k=1}^{K}\frac{\left[h_{3D_i}(k) - h_{3D_j}(k)\right]^2}{h_{3D_i}(k) + h_{3D_j}(k)} \tag{3.10}$$

where h_{3D_i} and h_{3D_j} are structural context histograms extracted from pyramidal shape maps of the gallery and probe sets and K is the number of histogram bins. For example, matched keypoints for two pyramidal SI maps of the same subject and different subjects are illustrated in figure 3.6b. Applying a histogram of the structural context of 3D pyramidal shape maps to extract local information has the advantage of improving matching scores under degradation conditions including pose, scale, rotation, translation, illumination, and expression variations. If there is a facial expression, the local areas such as nose, eyes, and so on that are invariant to expression changes cause our proposed method to be expression invariant throughout the matching process. In figure 3.6c, the differences between the numbers of matched keypoints on texture images, the SI map, and the pyramidal SI map are given. As the figure shows, applying the proposed pyramidal SI map can improve the matching score.

3.6.3 Score-level fusion

The final matching score from 2D and 3D face data can be fused in different ways. According to [122], the sum rule provides better results than other score fusion rules. In this work,

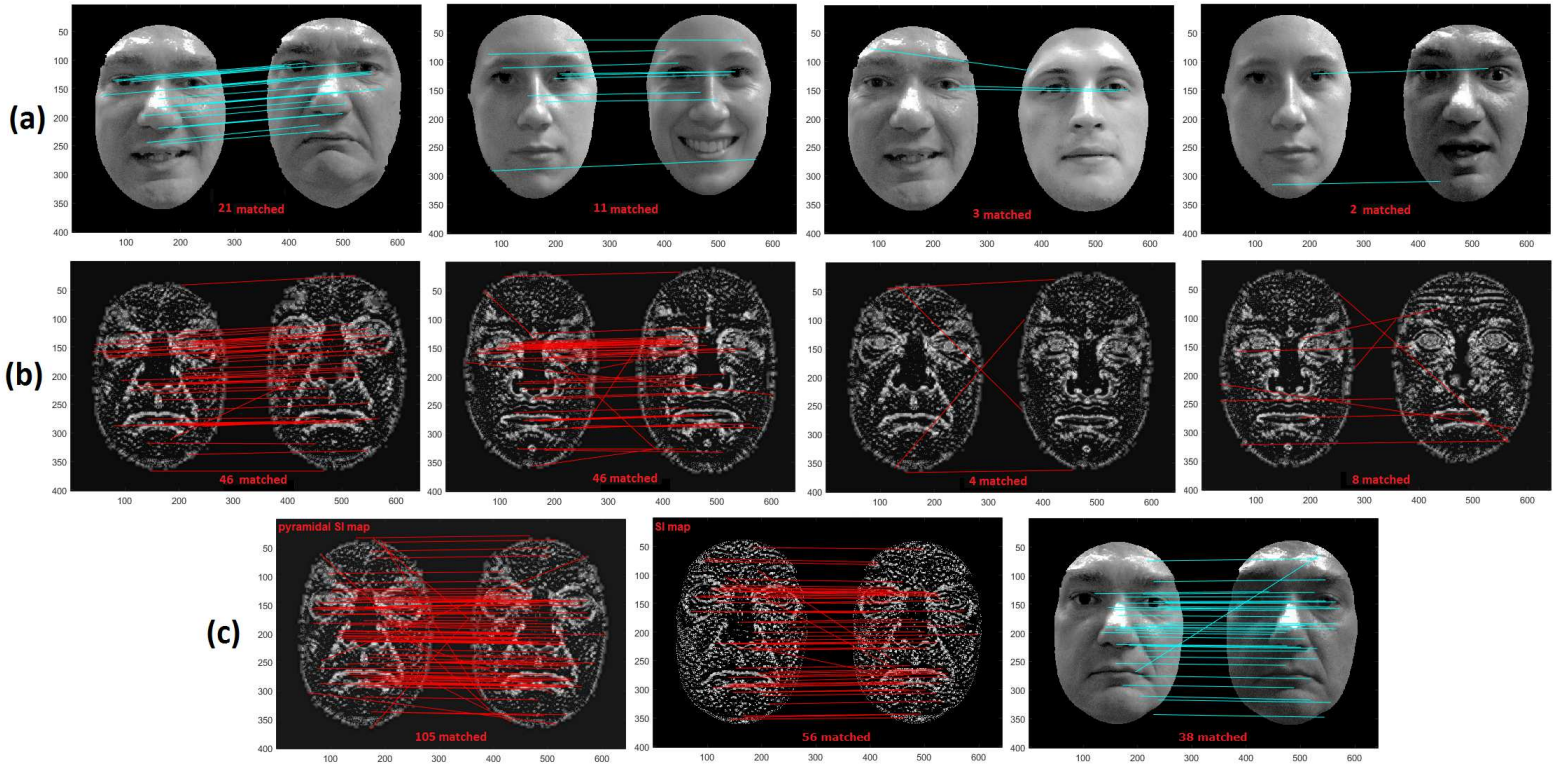

Figure 3.6 – Matched keypoints using a comparison of structural context for a) two texture images and b) pyramidal SI maps of the same subject and different subjects, c) comparison of number of matched keypoints from left to right for PSI map, SI map and texture image

the weighted sum based on 3.11 is applied to emphasize the 3D face features using a generalization of the sum rule, as it turns out that the histogram of structural context descriptors on pyramidal shape maps are more distinctive and reliable than the texture image.

$$s = k_{sc}s'_{sc} + k_{sm}s'_{sm} \tag{3.11}$$

In the above equation, s'_{sc} and s'_{sm} are normalized similarity scores obtained from 2D and 3D matching, respectively. To normalize scores, min-max normalization is applied to the scores produced by the 2D and 3D matches. Thus, we shift the minimum and maximum scores to 0 and 1, respectively, using 3.12. In this equation, for a set of matching scores $\{s_q\}$, $q = 1,2,,n$, the q^{th} element of each vector corresponds to the similarity between the probe and the q^{th} gallery face and s'_q is the normalized score. 2D and 3D matching scores are normalized to calculate final score s.

$$s'_q = \frac{s_q - min}{s_q - max} \tag{3.12}$$

As 3.11 shows, the overall similarity of the probe with the gallery is calculated using a confidence weighted sum rule, and k_{sc} and k_{sm} are the confidences for each individual similarity measure. To calculate these confidences based on the approach presented in [3], 3.13 is employed and can be conducted offline from the results obtained on training data or dynamically during online recognition. In this equation, $\bar{s}_q$, min, and min 2 are the mean value, first, and the second minimum of score vector, respectively.

$$k_q = \frac{\bar{s}_q - min}{\bar{s}_q - min2} \tag{3.13}$$

3.7 Experimental results

To evaluate the performance of the proposed method, FRGC database [16] that consists of large subjects with various facial expression and Bosphorus database [2] for more assess-

ments under expression, pose variations, and partial occlusions are utilized.

3.7.1 Experiments on FRGC database

In this section, the FRGC database [16] is applied to examine the proposed method. Different types of tests are employed to show the effectiveness of the algorithm. The database consists of FRGCv1 and FRGCv2 sets as a training and validation sets, respectively. Both sets were collected with a Minolta Vivid 3D scanner. The 3D data is given in a 640×480 grid. Each point in the grid has X, Y, and Z coordinates in millimetres and a valid flag. For each 3D scan, there is an accompanying 640×480 2D colour image. The first set, FRGCv1 (Spring 2003 session), includes 943 scans of 275 persons. All records are neutral and employed for training to determine the threshold values in both texture and shape modalities and confidences in score-level fusion. The second set, FRGCv2, (Fall 2003 and Spring 2004 sessions) is comprised of 4007 scans of 466 persons. The records including neutral and non-neutral images are applied for the validation phase according to the experimental protocol in [16]. The implemented system used MATLAB on a computer with an Intel Core i7 3.60 GHz CPU with 8 GB RAM. Among detected keypoints, some outliers are eliminated based on the approach presented in Section 5. By testing 1000 random matches, we could find our multimodal method achieves one full match in an average of 0.35 s. However, our main goal in this work is more on the verification accuracy, easy implementation, and robustness compared with the computational time. In the first experiments, from the validation set (4007 scans), the first sample of each subject with a neutral expression is considered as a gallery set (466 samples). The remaining scans, 3541 samples, are used as a probe set and consist of 1944 scans with neutral expression and 1597 scans with non-neutral expression. A verification result at 0.1% false acceptance rate (FAR) is usually reported in the literature as a general performance criterion. In shape modality, we evaluate the performance of the four proposed pyramidal shape descriptors for the neutral versus all test according to FRGC program [16]. A verification rate at 0.1% FAR is represented for

each shape descriptor in table 3.1. As the results show, the verification rate for the pyramidal SI is equal to 97.76%, which is the best 3D pyramidal shape map descriptor result compared with pyramidal curvedness, mean, and Gaussian curvatures. Based on this result, we continue our experiments only for the pyramidal SI as the most powerful 3D descriptor in the next experiments. ROC curves for texture and shape modalities and the results after score-level fusion for neutral versus neutral, neutral versus non-neutral, and neutral versus all tests according to FRGC program [16] are shown in figure 3.7a. According to this figure, at 0.1% FAR, the verification rate for our proposed 3D feature is 99.8% and 97.76% for neutral and all expressions respectively. The multimodal verification rate by means of scorelevel fusion is 99.9% and 99.3% for probes with neutral and all expressions respectively. Since the structural context descriptor is invariant against pose, scale, and illumination changes, and the proposed pyramidal SI works on shape information of 3D data and is invariant relative to expression variations in eyes and nose regions, the combination of these two descriptors can improve face recognition performance significantly. In the following experiments, we apply two protocols all versus all and ROC III according to the FRGC program [16]. In the all versus all verification experiment, all 4007 records are used as gallery and probe. A 4007×4007 full similarity matrix is obtained, and self matches are neglected. In the ROC III verification experiment, the gallery and probe records are from different sessions. The images taken in Fall 2003 are considered as the gallery set and the images taken in Spring 2004 form the probe set. Figure 3.7b illustrates the ROC curves for texture image, pyramidal SI map, and score-level fusion. From ROC curve for all versus all, the verification rate at 0.1% FAR for multimodal face recognition is 99%. This rate is 96.52% and 87.3% using the pyramidal SI map and texture image matching approach. For the ROC III experiment, which is the hardest experiment due to the time gap between gallery and probe records, the achieved results for verification rate at 0.1% FAR are 98.65%, 95.12%, and 84.2% for 2D and 3D fusion scores, 3D matching, and 2D matching, respectively. Tables 3.2 and 3.3 represent a comparison of our proposed method verification rates with related works. Some of the methods [22, 115, 7, 3, 116, 117] are

Table 3.1 – Verification rate for neutral versus all at 0.1% FAR for pyramidal shape map descriptors on two different databases

Pyramidal shape map	*SI%*	*C%*	*H%*	*K%*
on FRGCv2 [16]	**97.76**	94.54	91.15	88.73
on Bosphorus [2]	**93.46**	91.21	89.93	85.24

multimodal approaches that use 2D and 3D face data. Approaches presented in [81, 13] employ hybrid matching and feature- and score-level fusion, respectively. Compared with state-of-the-art, our proposed approach presents a higher verification rate at FAR 0.1% in all experiments.

For the identification experiment, the cumulative match characteristics curve is given in figure 3.8 for neutral versus all experiment. In this experiment, the neutral records are considered as the gallery and all of the records that are a combination of neutral and non-neutral expressions make up the probe set. The proposed multimodal approach has a 96.9% rank-1 identification rate and a 95.85% individual rate for 3D matching. In table 3.4, we compare the rank-1 identification rate for the neutral versus neutral and neutral versus all experiments with the state-of-the-art based on the results on FRGC database reported in the literature. The performance of our system has a higher rate for neutral versus neutral and non-neutral experiments. Although, the result for neutral versus all experiment is not the best in the table, but we can still state that the proposed algorithm achieves a high identification rate among similar methods.

3.7.2 Experiments on Bosphorus database

To further validate the effectiveness of the proposed method under degradation conditions including expression, pose variations, and partial occlusions, in this section the Bosphorus database [2] is employed. It contains 34 facial expressions (action units and six emotions), 13 pose variations (yaw, pitch, and cross-rotations), and 4 occlusions (eye with hand, mouth with hand, hair, and eyeglasses). It consists of 4666 records from 105 subjects. The 3D

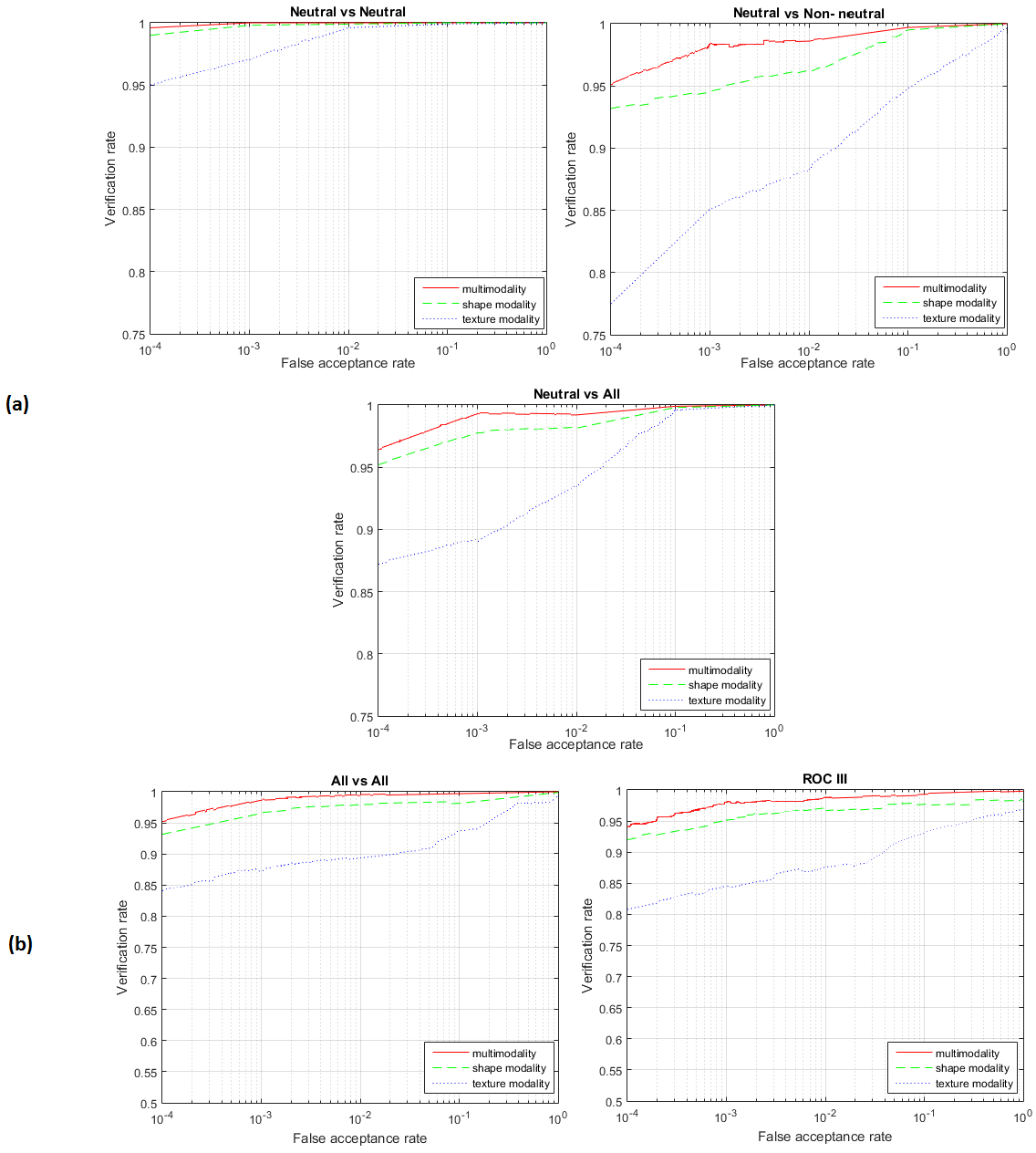

Figure 3.7 – ROC curves for texture and shape modalities and the results after score-level fusion a) Neutral vs Neutral, Neutral vs Non-neutral, and Neutral vs All experiments, b) All vs All, and ROC III experiments

point clouds were acquired with a structured-light technique, the Inspeck Mega Capturor II 3D scanner. In the experiments, we consider two different tests. First, the samples that

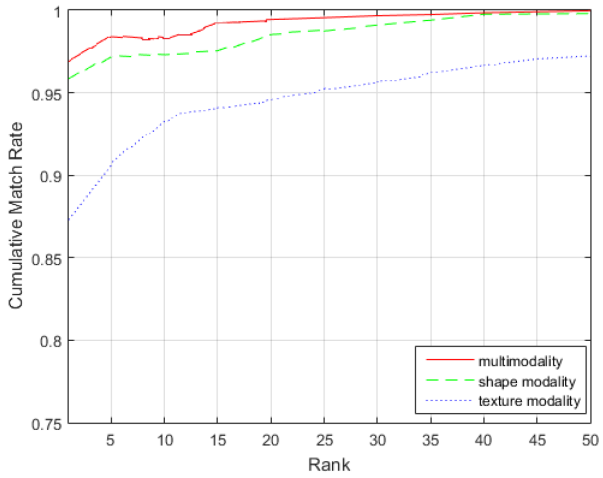

Figure 3.8 – Cumulative match characteristics curve

Table 3.2 – VR for Neutral versus neutral, Neutral versus nonneutral, and Neutral versus all at FAR = 0.1% on FRGCv2 database

Algorithm	Modality	Neutral versus neutral,%	Neutral versus non-neutral,%	Neutral versus all,%
[22]2007	2D+3D face (feature based and holistic)	99.74	98.31	99.3
[115]2011	2D+3D face (optimized weighted sum fusion)	99.9	97.1	98.9
[7]2015	2D+3D face (curvelet local features)	99.9	98	99.2
[3]2008	2D+3D face (keypoints and local features)	99.9	96.6	98.6
[116]2011	2D+3D face (resolution invariant local feature based)	99.5	92.9	96.3
[117]2012	2D+3D face (optimised data level fusion)	99.83	97.93	–
[81]2012	3D face (hybrid matching, local and holistic analysis)	99.6	97.2	98.4
[13]2013	3D face (fusion of local low-level features)	–	97.6	–
this work	2D+3D face (local descriptors, score-level fusion)	**99.9**	**98.5**	**99.3**

are nearly frontal consists of 3301 samples with facial expression or partial occlusion and second, all 4666 samples including pose variations are employed. In both experiments,

Table 3.3 – VR for All versus all and ROCIII at FAR = 0.1% on FRGCv2 database

Algorithm	Modality	All versus all,%	ROCIII,%
[7]2015	2D+3D face (curvelet local features)	-	97.8(FAR=0.1)
[111]2015	3D face, SI	-	95
[6]2014	3D face, keypoints	-	86.6
[81]2012	3D face (hybrid matching, local and holistic analysis)	94.2	95
[36]2010	3D face (local shape difference boosting)	98	98.1
[119]2010	3D face (curvature descriptors)	85.6	-
[108]2008	3D face (fusion of results)	94.8	93.2
[4]2013	3D face, keypints (meshSIFT)	79	77.2
[22]2007	2D+3D face (feature based and holistic)	-	86.6
this work	2D+3D face (local descriptors, score-level fusion)	**99**	**98.65**

Table 3.4 – Comparison of rank-1 identification rate on FRGCv2 database

Algorithm	Modality	Neutral versus neutral,%	Neutral versus all,%
[22]2007	2D+3D face (feature based and holistic)	99.02	96.2
[115]2011	2D+3D face (optimized weighted sum fusion)	99.6	98.0
[7]2015	2D+3D face (curvelet local features)	99.4	97.1
[3]2008	2D+3D face (keypoints and local features)	99.4	96.1
[116]2011	2D+3D face (resolution invariant local feature based)	99.4	96.2
[117]2012	2D+3D face (optimized data level fusion)	99.17	97.4
[81]2012	3D face (hybrid matching, local and holistic analysis)	99.2	97.6
[70]2015	3D face (feature level fusion)	–	96.3
this work	2D+3D face (local descriptors, score-level fusion)	**99.6**	96.9

a gallery set is constructed using the first neutral facial scan and the probe set is made

using the remaining samples. The verification rate of 3D modality on Bosphorus database at 0.1% FAR for each shape descriptor is shown in table 3.1 which the highest rate is for pyramidal SI map. Using this map, our multimodal method achieved the verification rate VR = 98.4% at FAR = 0.1% and rank-1 recognition rate RR = 97.2% for the first test and VR = 95.8% and RR = 94.5% for the second test. These results show that pose variation can affect the proposed method performance. Verification rate at FAR = 0.1% for second test compared with the multimodal approach [7] with VR= 91%, is 4.8% higher. The rank-1 recognition rate for the second test compared with feature-based fusion method [5] with RR = 94.1% is 0.4% higher. The fusion-based method in [70] achieved RR = 96.56% with high computational cost for the second test. For the first test, the rank-1 recognition rate of our method is 0.2% higher than the method [81] with RR = 97%. The results show that our algorithm achieves high performance on the Bosphorus database under facial expression and occlusion.

3.8 Conclusion

In this chapter, we proposed a novel local descriptor based multimodal algorithm for face recognition. Pyramidal shape map descriptors were proposed and applied to extract discriminative features from 3D data. Histograms of structural context were used in both 2D and 3D matching processes. Score-level fusion improves the final score efficiently. The proposed method is scale, translation, rotation, and expression invariant due to the use of SIFT keypoints, structural context, and pyramidal shape map descriptors. Experimental results on the most challenging 3D databases, FRGCv2 and Bosphorus illustrate high performance of the proposed approach. In the future work, feature extraction from 3D data without 2D feature support that is more robust against facial challenges will be studied.

Chapter 4

Multiscale Depth Local Derivative Pattern for Sparse Representation Based 3D Face Recognition

3D face recognition is a popular research area due to its vast application in biometrics and security. Local feature-based methods gain importance in the recent years for their robustness under degradation conditions. In this chapter, a novel high-order local pattern descriptor in combination with sparse representation based classifier (SRC) is proposed for expression robust 3D face recognition. 3D point clouds are converted to depth maps after pre-processing. Multidirectional derivatives are applied in spatial space to encode the depth maps based on the local derivative pattern (LDP) scheme. Directional pattern features are calculated according to local derivative variations. Since LDP computes spatial relationship of neighbors in a local region, it extracts distinct information from the depth map. Multiscale depth-LDP is presented as a novel descriptor for 3D face recognition. The descriptor is employed along with the SRC to increase the range data distinctiveness. A histogram on the derivative pattern creates a spatial feature descriptor that represents the distinctive micro-patterns from 3D data. We evaluate the proposed algorithm on two famous 3D face databases, FRGCv2 and Bosphorus. The experimental results demonstrate that the proposed approach achieves acceptable performance under facial expression.

4.1 Introduction

For many years 2D face recognition has been studied among researchers as an important and popular biometrics. However, degradation conditions such as illumination and pose variations have influenced on 2D face recognition system performance. To overcome these limitations, 3D face data that contains spatial information has attracted researchers attention.

Applying efficient descriptors in 2D face recognition on depth images helps researchers to achieve high performance in 3D dimensionality like Gabor wavelet [123] and LBP [64]. Huang et al. [64] proposed feature-based method using shape index (SI) and local binary pattern (LBP) for 3D facial surface representation. LBP is considered as a simple and most efficient local 2D face descriptor that is first proposed by Ojala et al. [42]. Recently, LBP has been applied by researchers as an effective local descriptor in 3D face area. Multi-scale extended LBP [81] that is a facial surface descriptor extracts local shape changes and applies SIFT-based matching for face recognition. In [98], depth and normal information of 3D data are extracted and encoded using LBP to create a face descriptor. The surface normal that determines a surface orientation at each point and includes local shape information also is applied by Li et al. [15] for feature-based 3D face recognition. Local normal pattern inspired by LBP is used to describe shape information and extended as a multiscale and multicomponent descriptor to improve the recognition system performance. To handle facial expression, they applied a weighted sparse representation-based classifier (W-SRC). The whole face is divided into local patches and local normal-based features are extracted and used in the training step to learn weights. The W-SRC is also employed in [124] along with region-based extended LBP descriptor for 3D face recognition.

Sparse representation [125] which is a subspace algorithm can be used as a feature representation method to extract more distinct feature and dimension reduction for depth images. The sparse regression model is proposed in [126] to embed the facial descriptors into the low dimensional matrix and handle occlusions and hair covering. In [102],

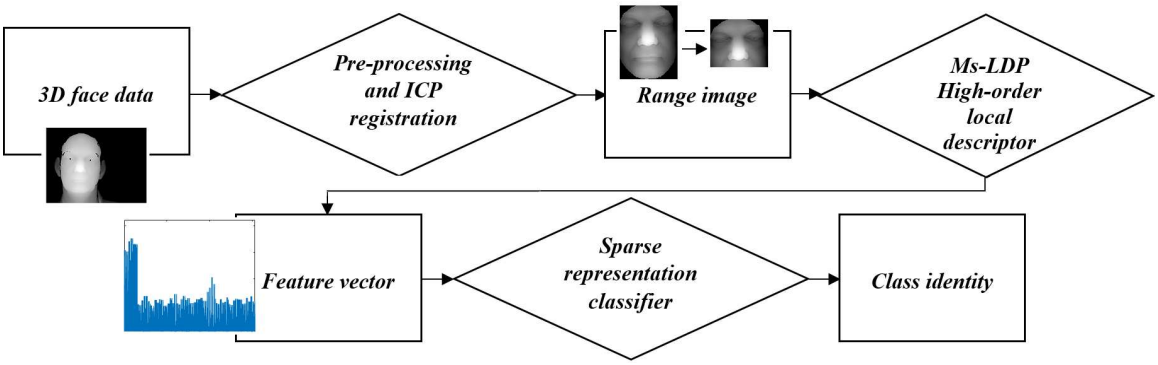

Figure 4.1 – The framework of the proposed method

authors apply sparse representation framework in combination with feature pooling and ranking scheme for low-level features. A local descriptor called local shape pattern (LSP) is presented by Huang et al. [127] to extract both differential structure and orientation information from 3D face data. They applied SRC to classify the local shape features.

In this chapter, we propose multiscale depth local derivative pattern (MsDLDP) for 3D face recognition. Unlike LBP that is a non-directional first-order local pattern, LDP captures the change of derivative directions among local neighbors. High-order LDP performs better results for extracting more discriminative features compared to LBP. In our proposed algorithm, we apply learning-based approach using sparse representation (SR) to select prominent features and boost recognition rate. Similar to LBP, LDP is also modeled using a histogram of the extracted micro patterns. In the proposed approach, the histogram of MsDLDP is fed into SR classifier to do recognition task. The overview of the proposed approach has been presented in figure 4.1.

The remaining part of this work is presented as follows: in section 2, pre-processing of 3D face data is explained. Section 3 provides the proposed method that consists of feature extraction, MsDLDP descriptor, and classification using SR. Experimental results including the proposed algorithm performance and comparison with state-of-the-art is presented in section 4, and section 5 describes conclusion and future work.

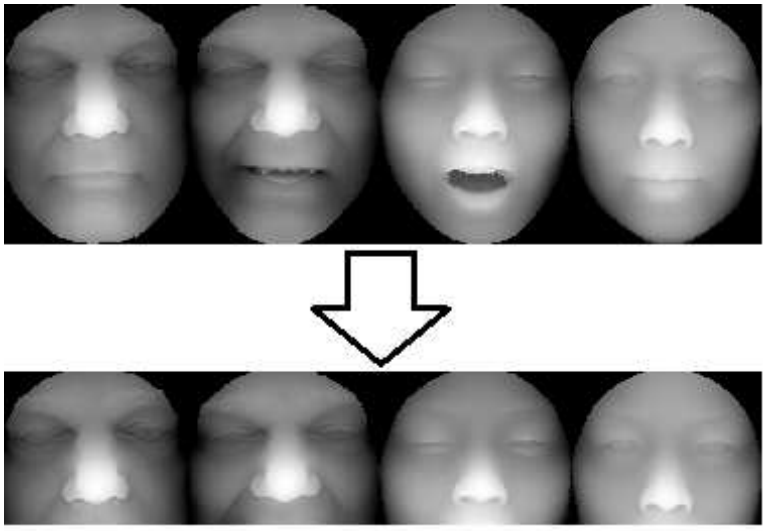

Figure 4.2 – Illustration of samples of the face pre-processing on the FRGCv2 database

4.2 Pre-processing

The output of the 3D capturing device is noisy and needs to be smoothed. In the pre-processing part, the 3D face scans are processed to smooth noise, remove spikes, and fill holes [1]. The region of interest (ROI) extraction is next part that is performed using nose tip detection.

In this chapter, we apply 3D face pre-processing tool developed by Szepticki et al. [128]. The median filter is used to remove spikes and noises. Hole filling is done using the square surface fitting. The curvature-based method is employed to detect nose tip and ROI extraction.

We apply iterative closest point (ICP) algorithm [129] to correct pose by considering five frontal scans with neutral expression from each database as models. We resize each pre-processed range image into 120×96 for next steps. To handle facial expression, we consider the rigid (nose) and semi-rigid (eye-forehead and cheek) areas and exclude the most impressed area by the expression, non-rigid area (mouth). The pre-processed faces from FRGCv2 database [16] have been shown in figure 4.2.

4.3 Proposed method

4.3.1 Multiscale depth local derivative pattern descriptor

Depth-LDP descriptor

Local derivative pattern on depth map that is a high-order multidirectional derivative texture pattern descriptor is proposed in this section for 3D face. LDP first is introduced by Zhang et al. as a texture pattern descriptor for 2D images [130]. It can be considered as a denoising function for its special binarization function.

To calculate LDP for a depth image $D(P)$, the first-order derivative $D'(P)$ for different directions including $0^o, 45^o, 90^o$, and 135^o is calculated using the following equations

$$D'_{0^o}(P_0) = D(P_0) - D(P_4) \tag{4.1}$$

$$D'_{45^o}(P_0) = D(P_0) - D(P_3) \tag{4.2}$$

$$D'_{90^o}(P_0) = D(P_0) - D(P_2) \tag{4.3}$$

$$D'_{135^o}(P_0) = D(P_0) - D(P_1) \tag{4.4}$$

where, P_0 is a point in $D(P)$ and $P_i, i = 1, ..., 8$ is the neighboring point around the P_0 as figure 4.3 shows.

To compute second-order directional LDP with direction α at P_0, the following equation is applied.

$$\begin{aligned} DLDP^2_\alpha(P_0) = (&f(D'_\alpha(P_0), D'_\alpha(P_1)), f(D'_\alpha(P_0), D'_\alpha(P_2)) \\ &, ..., f(D'_\alpha(P_0), D'_\alpha(P_8))) \end{aligned} \tag{4.5}$$

A binary coding function f is used to determine local pattern transition types. The consistency of two neighboring derivatives is described using f defined as

$$f(D'_\alpha(P_0), D'_\alpha(P_i)) = \begin{cases} 0, & \text{if } D'_\alpha(P_0).D'_\alpha(P_i) > 0 \\ 1, & \text{if } D'_\alpha(P_0).D'_\alpha(P_i) \leq 0 \end{cases} \tag{4.6}$$

The encoding system is applied on the binary derivative calculations to make the integer value of the LDP descriptor. In our proposed method, according to above explanation, each pixel in depth map is assigned an integer value at specific direction α. Each depth map is divided into some local patches. The statistical distribution of the calculated features in the local regions is computed and presented using a histogram. The length of each histogram is 2^m which m is the number of the neighbors around the center point P_0. The calculated histograms from each local patch are concatenated together to make the histogram in each direction. Final descriptor is created using histogram concatenation of different directions (see figure 4.4).

Multiscale approach

Like LBP, LDP can be extended with different local neighborhood sizes for different scales. A set of sampling points around the central point P_0 is considered as the local neighborhood. The arrangement of the sampling points is defined using a various number of the points and radius (P,R). Figure 4.3 illustrates different LDP neighborhoods. MsLDP is defined by changing the value of radius R. This scheme for LBP, MsLBP, firstly is proposed by Ojala et al. for texture classification [42] and applied for 2D face recognition by Chan et al. [131]. Later, Huang et al. [64] applied it for 3D face recognition. In this work, we propose a multiscale strategy for depth-LDP calculation that is quite a different and new presentation of LDP for 3D face recognition. The local derivative pattern at different radiuses computes local shape variations and extracts highlight details.

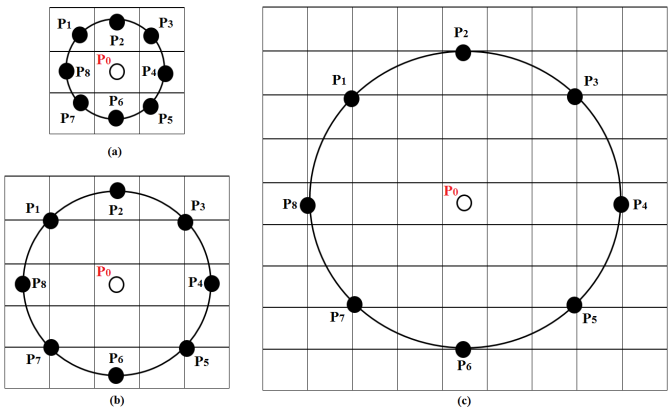

Figure 4.3 – Eight neighborhood around P_0 with different R, (a) R=1, (b) R=2, (c) R=3

n^{th}-order MsDLDP descriptor

The higher-order derivatives of the MsDLDP is computed by applying equations 4.1 through 4.4 on MsDLDP iteratively and using equation 4.6 for binarization. In this way, the n^{th}-order derivative for our proposed descriptor is calculated as follows:

$$MsDLDP^{n}_{\alpha}(P_0) = f(D^{n-1}_{\alpha,R}(P_0), D^{n-1}_{\alpha,R}(P_i)), i = 1,...,8 \tag{4.7}$$

where R is the different values for the radius to generate the multiscale descriptor.

4.3.2 Sparse Representation-based classifier

The sparse representation classifier firstly introduced by Wright et al. [125] for 2D face recognition. They consider the problem of recognizing of the frontal faces under varying expression and lighting which can be addressed using sparse signal representation. L1-minimization is used to compute a sparse representation as a general classification algorithm. This framework provides the insight that if the sparsity could be harnessed properly, the performance of the classification would be improved. Base on this representation, for a frontal test sample, the sparsity of the coefficient vector is high except for the same class samples. These coefficients for the ones from other classes are zero or close to zero.

We apply the above framework for 3D face scans by considering the probe face as a

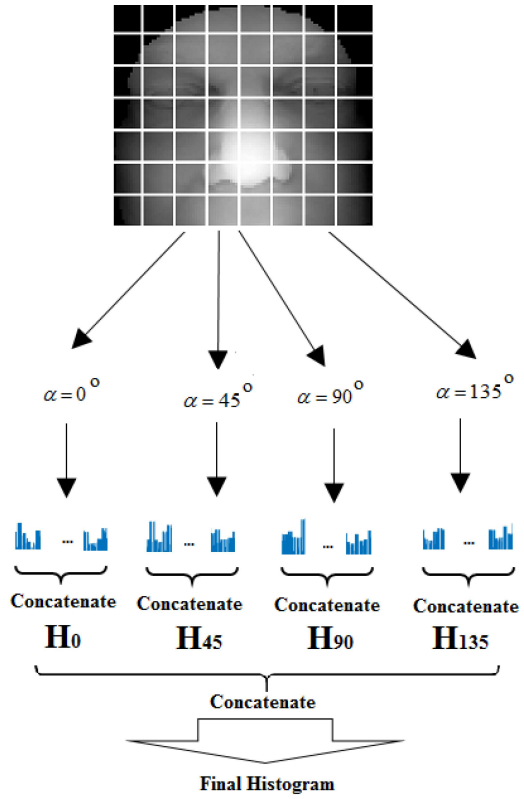

Figure 4.4 – Construction of depth-LDP descriptor

sparse linear combination of gallery samples. Given n_i training 3D face samples of the class i, $A_i = [v_{i,1}, v_{i,2}, ..., v_{i,n_i}] \in R^{m \times n_i}$, any probe sample from the same class is defined using the linear relation of the training samples in class i based on the following equation

$$y = \beta_{i,1} v_{i,1} + \beta_{i,2} v_{i,2} + ... + \beta_{i,n_i} v_{i,n_i} \tag{4.8}$$

where $\beta_{i,j} \in R$, $j = 1, 2, ..., n_i$. A new matrix A is defined for all training samples by concatenating of the n samples from i different 3D face classes. In this way, the linear representation of the probe sample can be defined as follows:

$$y = Ax_0 \in R^m,$$
$$x_0 = [0,...,0,\beta_{i,1},\beta_{i,2},...,\beta_{i,n_i},0,...,0]^T \in R^n \tag{4.9}$$

where x_0 is a coefficient vector with entries equal to zero or close to zero except those related to the subject of y. It is obvious that when $m > n$, the system based on the equation 4.9 is over-determined and the vector x_0 can easily be found as its unique solution. However, in 3D face recognition application, it is worth noting that for each subject there is only one sample in the gallery as training sample which is based on the most common setting in 3D face recognition systems. Consequently, in 3D face recognition applications the system $y = Ax_0$ is typically under-determined that its solution is not unique. For the frontally aligned faces, the expression variation problem is another challenge that we handle it in our proposed method by considering the rigid, and semi-rigid parts of the face as it is explained in section 2. To solve the sparse vector x the following solution using the minimum $l^0 - norm$ according to [125] is applied as follows:

$$\hat{x_0} = argmin||x||_0 \qquad s.t.||Ax - y||_2 \leq \varepsilon \tag{4.10}$$

where $\varepsilon \in R^n$ and represents a deviation vector. For sparse x_0 the above equation can be solved using solving the problem of L_1-Norm [125] and reconstruction residuals $r_i(y)$ calculation as follows:

$$\hat{x_2} = argmin||x||_2 \qquad s.t.||Ax - y||_2 \leq \varepsilon \tag{4.11}$$

$$r_i(y) = ||y - A\delta_i(\hat{x_1})||_2^2 \tag{4.12}$$

where δ_i represents a characteristic function that is used to select the coefficient related to the i^{th} sample of the gallery. Consequently, the identity of the probe y is determined with the index of minimal $r_i(y)$.

4.4 Experimental results

In this section, we evaluate the proposed approach that consists of a new facial local descriptor along with sparse representation classifier. The following databases are used for the comprehensive evaluation of the proposed method under expression variations.

4.4.1 FRGC DB

The FRGCv2 [16] database consists of 4007 texture and 3D face scans under different facial expression from 466 persons. It is the largest set of the 3D face database that has been used in literature as the benchmark for 3D face recognition algorithms evaluation. The face acquisition system is the Minolta Vivid 900 scanner. The face scans are in controlled lighting and pose and they are under facial expressions such as happiness and surprise. In the experiments, the gallery consists of 3D face scans with neutral expression from each subject that are 466 samples to make the dictionary A of the SRC after applying MsDLDP descriptor of these samples. The remaining samples including 3541 3D face scans make up probe samples, y in equation 4.8.

4.4.2 Bosphorus DB

To further evaluation of the proposed algorithm under facial expression variations, in this section, we apply the Bosphorus database [2] that compromises of 4666 texture and 3D records from 105 persons. This database contains 34 facial expressions including action units and 6 emotions, 13 different pose variations that consist of the pitch, yaw, and cross rotations, and 4 occlusions including hair, eye glasses, eye, and mouth with the hand. The

hardware device used to capture 3D face scans is the Inspeck Mega Capturor II 3D scanner. For the experiments using the Bosphorus database, we apply same approach for the pre-processing used for the FRGC database. In this experiment, we consider nearly frontal faces with expression changes and partial occlusions. These scans consist of 3301 samples. A gallery set (105 scans) is made up using the first sample with the neutral expression and the remaining samples make up the probe set (3196 scans).

4.4.3 Experiments

To compare different orders of the proposed multiscale local descriptor, the recognition rate of different orders using SRC has been reported for neutral vs. all experiment based on the experimental protocols presented in [16]. Increasing order of the local pattern can improve the recognition results for second and third-order descriptors based on the results presented in figure 4.5. Since the accuracy of the recognition system using fourth-order has been decreased, it means that not only increasing the order to four does not add more information but also causes to convert the facial image to the noisy data and destroy the recognition rate.

In the next experiment, we evaluate the effectiveness of sparse representation-based classifier. The Chi-square distance is used in the literature as a popular similarity measurement for histogram-based descriptors like LBP, LDP, and etc.[132]. To show the effectiveness of sparse-based classifier, we compare the recognition rate of SRC and Chi-square based classifier using the same facial descriptor with different orders. Tables 4.1 and 4.2 present the rank-one recognition rate (RR1) for neutral vs. all experiment on FRGCv2 and Bosphorus databases respectively. From the tables, it is obvious that the SRC outperforms the performance of Chi-square classifier and shows applying sparse-based classifier is effective along with local derivative pattern descriptor.

To evaluate the effectiveness of the proposed approach under facial expression, we test our algorithm on two sets from FRGCv2 database by dividing the probe samples into two

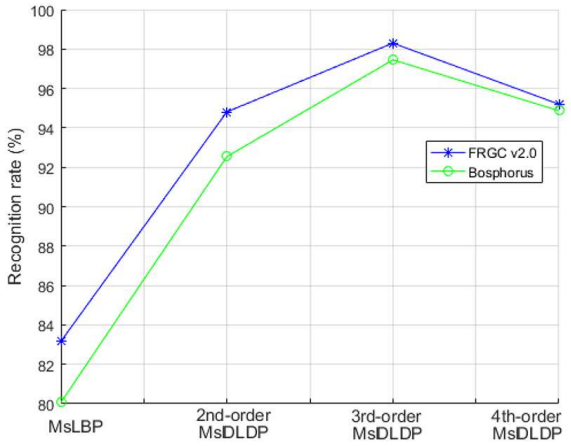

Figure 4.5 – Comparison of rank-one recognition rate of MsDLDP descriptor on FRGCv2 and Bosphorus databases for different orders

Table 4.1 – The comparison of RR1 for two different classifiers on the FRGCv2 database

Local descriptors	RR1 (Chi-square)	RR1 (SR)
MsLBP	75.34%	83.16%
MsDLDP(2nd-order)	82.7%	94.81%
MsDLDP(3rd-order)	**88.46**%	**98.3**%
MsDLDP(4th-order)	83.6%	95.2%

Table 4.2 – The comparison of RR1 for two different classifiers on the Bosphorus database

Local descriptors	RR1 (Chi-square)	RR1 (SR)
MsLBP	72.8%	80.06%
MsDLDP(2nd-order)	79.1%	92.54%
MsDLDP(3rd-order)	**85.9**%	**97.45**%
MsDLDP(4th-order)	82.41%	94.87%

different sets including neutral and non-neutral facial scans. In this experiment, we consider the third-order descriptor that is the most effective one based on the experimental results. Table 4.3 reports the performance of the system under facial expression and compares the obtained results with the state-of-the-art. As the results illustrate our proposed method is robust under expression by applying the local derivative pattern in the multiscale scheme that extracts discriminant enough information and as we exclude the non-rigid parts in the pre-processing step.

Table 4.3 – The performance of the proposed method under facial expression on the FRGCv2 database

Methods	RR1 (Neutral)	RR1 (Non-neutral)
Huang et al.[64]	99.1%	92.5%
Li et al.[15]	98%	94.2%
MsDLDP(3rd-order)+SRC	**99.3**%	**97.1**%

Table 4.4 – The performance comparison with LBP-based methods on the FRGCv2 and Bosphorus databases

Methods	RR1 (FRGCv2 DB)	RR1 (Bosphorus DB)
Huang et al.2012[81]	97.6%	97%
Tang et al.2013[98]	94.89%	-
Li et al.2014[15]	96.3%	95.4%
Lv et al.2015[124]	97.8%	-
This work	**98.3**%	**97.45**%

In the next experiment, the proposed approach is compared with the state-of-the-art 3D face recognition approaches that belong to LBP-based category method. We evaluate our algorithm on both databases for neutral vs. all and compare with other methods under the same experimental conditions. The performance comparison is reported in table 4.4. From the results presented in the table, we can find that our proposed descriptor in combination with sparse representation-based classifier outperforms state-of-the-art. Among different local pattern-based descriptor in the table, the local derivative pattern can extract more distinct features since it works based on the spatial relationship of the neighbor points in the local region.

4.5 Conclusion

In this chapter, a new facial descriptor called high-order multiscale depth local derivative pattern (MsDLDP) has been proposed. The descriptor contains more spatial information compared to local binary pattern (LBP), since it encodes the various distinct spatial relationship in a local region. The proposed multiscale strategy provides more discriminative in-

formation and represents local shape features more comprehensively. In addition, to select most distinct features and improve recognition performance sparse representation-based classifier has been employed. The experimental results illustrate that the SRC is more efficient than distance-based classifier. The algorithm can handle expression variations properly since we exclude the expression sensitive non-rigid areas in the pre-processing step. The presented algorithm is robust under facial expression for aligned nearly frontal faces. However, it is sensitive to pose variations.

In the future, we will extend our proposed descriptor on different shape maps and Gabor features. The weighted classifier to handle expression is another research direction that we will work on it to apply the whole face including non-rigid areas. Also, working on the recognition system robustness under pose variations is another plan to extend our work to face with this challenge.

Chapter 5

Weighted Extreme Sparse Classifier and Local Normal Derivative Pattern for 3D Face Recognition

A novel weighted hybrid classifier and high-order local normal derivative pattern descriptor is proposed for 3D face recognition. Local derivative pattern captures detailed information based on local derivative variation in different directions. LDP is computed on three normal maps in x, y, and z directions and different scales. Surface normal captures the orientation of a surface at each point of 3D data. Compared to depth, more informative local shape information is extracted using surface normal. The n^{th}-order LDP on the surface normal is proposed to encode more detailed features from $(n - 1)^{th}$-order local derivative direction variations. An extreme learning machine-based autoencoder using multilayer network structure is employed to select more discriminant features and provide faster training speed. A weighted hybrid framework is proposed to handle facial challenges by a combination of ELM and sparse representation classifier. The speed advantage of ELM and the accuracy advantage of SRC in a weighted scheme is used to enhance the performance of the recognition system. Experimental results on four famous 3D face databases illustrate the generalization and effectiveness of the proposed method in both computational cost and recognition accuracy.

5.1 Introduction

Generally, there are two critical factors related to any face recognition system. First, facial feature extraction which needs to be not sensitive under various challenges and distinctive for different subjects. Second, the design of a classifier with distinguishing capability between genuine and imposter samples.

In this work, we propose a novel local derivative descriptor to robustly recognize person's identity. There are some characteristics for an effective descriptor including high ability to differentiate between classes, low intraclass variations, and low computational complexity. Pose correction is performed using rigid-ICP [133] algorithm to extract pose invariant features. However, under extreme pose variation, the feature extraction may fail due to self-occlusion. To overcome the mentioned problem where some parts of the face are not visible, we propose a weighted hybrid classifier which combines sparse representation and extreme learning machine as a powerful classifier to manage noisy and incomplete data with fast learning speed.

The main contributions of this work are as follows:

A high-order descriptor called multiscale local normal derivative pattern, MsLNDP, is proposed which is able to robustly represent facial images under expression and pose variation. The proposed descriptor works on surface normals in x, y, and z directions and applies score level fusion in a multidirection scheme for a final decision. An ELM-based dimension reduction method is employed to extract distinct efficient features. A learning-based framework is employed to compute local patches weights of 3D facial surfaces to make discriminant features that are robust under facial challenges. A novel hybrid classifier called weighted extreme sparse classifier, WESC, is proposed which consists of two steps: first, learning of an ELM network and adopting a discriminant criteria to decide about ELM output reliability is performed. Second, in the case of unreliable output, the test image is fed into sparse representation classifier. By extracting sub-dictionary from ELM output the computational cost of the SRC can be reduced. To the best of our knowledge, no high-order

local pattern has been applied for 3D face representation and the combination of ELM and SR is the first attempt in the literature to recognize the 3D face.

5.1.1 Local Feature-based Methods

Face recognition system consists of different modules that one of them is feature extraction. It has an important role to handle degradation conditions and improve system efficiency. Feature extractors are divided into two different categories: global and local features [1]. There are several reasons that make local features more promising than global ones including their robustness under facial expression, occlusion, and missing parts [29]. A comprehensive literature survey on local feature methods for 3D face recognition can be found in [23]. A summary of local feature methods categorization is shown in figure 5.1. As the figure illustrates local features are divided into three different categories consist of keypoints, curves, and local surface features.

3D keypoints are computed using geometrical information of the surface to define shape saliency [67]. Since these methods use a large number of interest points it causes to increase computational complexity. One of the first methods for 3D keypoints is proposed by Mian et al. [3] using principle component analysis and scale invariant feature transform. Some methods extract keypoints directly from mesh data to handle large pose variations or occlusions like meshSIFT [4], meshDoG [6], and meshCurvelet [7]. Some of them are applied to different facial maps. For instance, SIFT-like keypoints on curvature maps using hybrid scheme have been proposed in [66]. Landmarks are another kind of keypoints extracted based on the anatomical studies of the face. Their main disadvantage is their sparsity.

Curves contain rich geometrical information by capturing shape features from different parts of the 3D face. Compared with keypoints, curves present less sparse features from the facial surface. They are divided into contours such as level curves including iso-depth [9] and iso-geodesic curves [82] and profiles like radial curves [10]. Nose tip is used as a reference point in most of the curve-based methods. Since nose region is rigid, these meth-

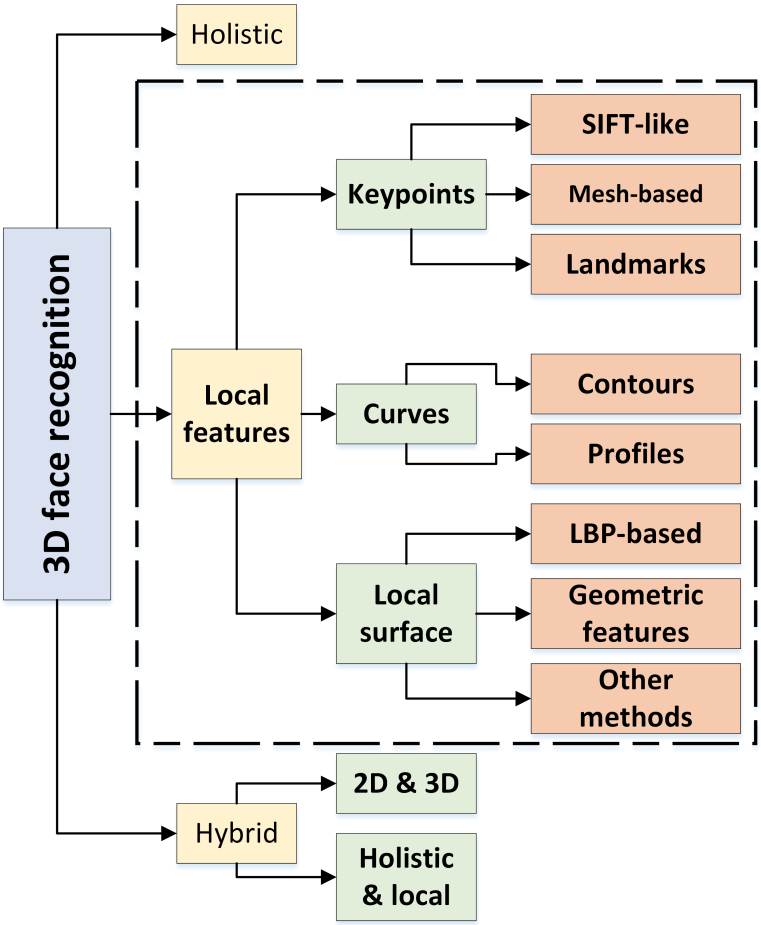

Figure 5.1 – Local feature methods categorization

ods can handle facial expression effectively. Correct nose tip detection can be influenced by hair covering, large pose variation, and missing data. Consequently, exact curve extraction and then system performance can be affected. Recently, nasal patches and curves have been applied in [95] for expression robust 3D face recognition. Most of the methods in this category have been proposed to handle facial expression [23].

Local geometric information is extracted from different patches or regions of the facial surface. Consequently, they can handle facial expression effectively. Local surface methods are divided into LBP-based, geometric features, and other methods [23]. LBP is one of the most effective local surface patterns which is initially introduced by Ojala et al. [42] as a texture descriptor. It has been applied in fusion scheme with intensity image [96] for

3D+2D based face recognition, then used for 3D face recognition such as 3DLBP [97], multiscale extended LBP [81], local normal pattern [15], and mesh-LBP [12]. Low-level geometric features which consist of distances and angles between 3D mesh vertices have been proposed by Lei et al. [13]. Covariance matrices of descriptors [103] to compute spatial and geometric properties of a region is another example for this category. These geometric features-based algorithms are not able to handle occlusion and missing data. Since the local features inspired by LBP demonstrate efficient performance on 3D face recognition [15], [81], [12] in terms of competitive performance and computational efficiency, we propose a new descriptor by encoding spatial relationship in a local region for different directions.

5.1.2 ELM-based Methods

One of the most efficient machine learning algorithms for pattern recognition and multiclass classification is the extreme learning machine. Compared to support vector machine, ELM needs milder optimization constraints, that improves learning speed with considerable performance [134]. 2D face recognition based on ELM has been discussed in many types of research such as [135, 136, 137, 138]. The basic and regularized ELM are adopted for face recognition in [135]. ELM is employed by Choi et al. [136] to learn face local patches sequentially. Basic ELM also is used in [137] with tensor subspace analysis for face representation. Baradarani et al. [138] apply ELM for face recognition to solve shortcoming of conventional methods like SVM and neural network including slow learning speed and poor computational scalability.

5.1.3 Sparse-based Methods

Sparse representation classifier has shown encouraging results in image classification [139, 140]. It was proposed by Wright et al. [125] for 2D face recognition under expression, occlusion, and illumination. Sparse representation coefficients of a query sample are es-

timated using a dictionary whose basis atoms are features extracted from gallery samples. There are several advantages to the methods based on the sparse representation. Finding the only one optimal matching by solving L-norm minimization problem is effective for face recognition applications. Since SR applied error separation property, the presence of some irrelevant features can not affect the system performance. However, it only causes to increase computational cost. In the SR framework, the number of persons in the gallery will enhance the sparsity and will not destroy the performance like the conventional methods. In this way, it is good for 3D face recognition in which there is one sample per person in the gallery. SR has been applied to geometrical features after feature ranking based on the fisher linear discriminant analysis in [141] for 3D face recognition. It also is employed to construct a patch-based point correspondence model of 3D faces [142] and to analyze feature representation [126]. Moreover, SRC is used to classify local shape patterns by Huang et al. [127]. SR classification method has been applied to local patterns extracted from depth map in [143]. To handle facial expression non-rigid regions of the face that are very sensitive under expression variations are excluded. Spherical sparse representation is proposed in [144] for dimension reduction on depth images for 3D face recognition. Face matching is done using the sparse comparison of facial features in [90]. SR framework on 3D faces is employed in [102] using low-level geometric features after feature pooling and ranking. Although SR presents superior robustness to occlusions, it cannot handle facial expression directly. Facial expression may cover entire facial area, whatever occlusion only occur some parts of the face. Weighted SR classifier is proposed by Li et al. [15] to handle facial expression variations using local normal histograms. However, they did not address the other challenges related to the recognition system.

According to the simple implementation, high learning speed and good generalization performance of the ELM and the efficiency of SRC in term of accuracy we apply a hybrid scheme on local features for 3D face recognition to handle various facial challenges. To the best of our knowledge, this approach is the first research on 3D face recognition that makes sub-dictionary for sparse representation based on the results of ELM method to improve

the recognition system accuracy and computational cost.

The rest of this chapter is organized as follows. Section 2 provides details on the proposed MsLNDP descriptor. The proposed weighted hybrid classifier, WESC, is described in section 3. Section 4 presents the experimental setting and recognition results to verify the efficiency and effectiveness of the proposed algorithm, while section 5 concludes this work.

5.2 Proposed descriptor

We propose a novel feature extraction on surface normals that provides more distinct information compared to the depth map. This work significantly extends our previous work [145] in which local derivative pattern on normal components was extracted and matched using a simple histogram intersection to handle only facial expression. However, in this chapter, we improve the distinctiveness of the descriptor by using multiscale scheme and auto-encoder for effective feature selection.

5.2.1 Surface Normal

This work is inspired by recent algorithms [38, 95, 15] in which surface normals have been applied for 3D face recognition. For a set of n points $P = \{p_1, p_2, ..., p_n\}$ of 3D point cloud $p_i \in R^3$, the data matrix is $P = [p_1, p_2, ..., p_n]^T$ where $p_i = [p_{ix}, p_{iy}, p_{iz}]^T$ based on the 3D coordinates of the points. A normal vector $n_i = [n_{ix}, n_{iy}, n_{iz}]^T$ is defined for every point p_i using a set of k neighbor points $Q_i = \{q_{i1}, q_{i2}, ..., q_{ik}\}$, $q_{ij} \in P$, $q_{ij} \neq p_i$. The neighbor matrix Q_i and the augmented one Q_i^+ including all neighbors and the central point p_i are defined as follows

$$Q_i = [q_{i1}, q_{i2}, ..., q_{ik}]^T, Q_i^+ = [p_i, q_{i1}, q_{i2}, ..., q_{ik}]^T \tag{5.1}$$

Normal estimation approaches are divided into optimization and averaging method [146]. In this work, we adopt optimization method since it can be applied to both 3D point clouds and mesh 3D data types [15]. In this method, normal vector n_i is calculated by solving the optimization problem $minA(p_i, Q_i, n_i)$ where A is a cost function to penalize a certain criteria which can be the distance of points to a local plane or the angle between normal and tangential vectors. (see figure 5.2). Based on the normal estimation, each 3D range image P with $m \times n \times 3$ data matrix is defined by three normal components in x, y, and z direction as follows

$$\begin{aligned} N(P) &= \left[n_{jk}^x, n_{jk}^y, n_{jk}^z\right], \\ & N_x = n_{jk}^x, N_y = n_{jk}^y, N_z = n_{jk}^z, \\ & 1 \leq j \leq m, 1 \leq k \leq n \end{aligned} \tag{5.2}$$

where $\|(n_{jk}^x, n_{jk}^y, n_{jk}^z)\|_2 = 1$.

For comparison between depth map and three normal components, figure 5.2 shows some samples from same and different subjects of face recognition grand challenge, FRGC database [16]. As it is obvious, normal components contain more information compared to the depth image.

5.2.2 Multiscale local normal derivative pattern

In this section, LDP is introduced and adopted on three normal components and depth map to create a novel descriptor in 3D. LDP first proposed by Zhang et al. [130] for 2D face recognition. Inspired by LBP as a gray-scale invariant texture descriptor, LDP works on high-order derivative variations. To compute LBP, a 3×3 neighborhood of each pixel is considered. The threshold function is applied to each central point and its neighbors (see

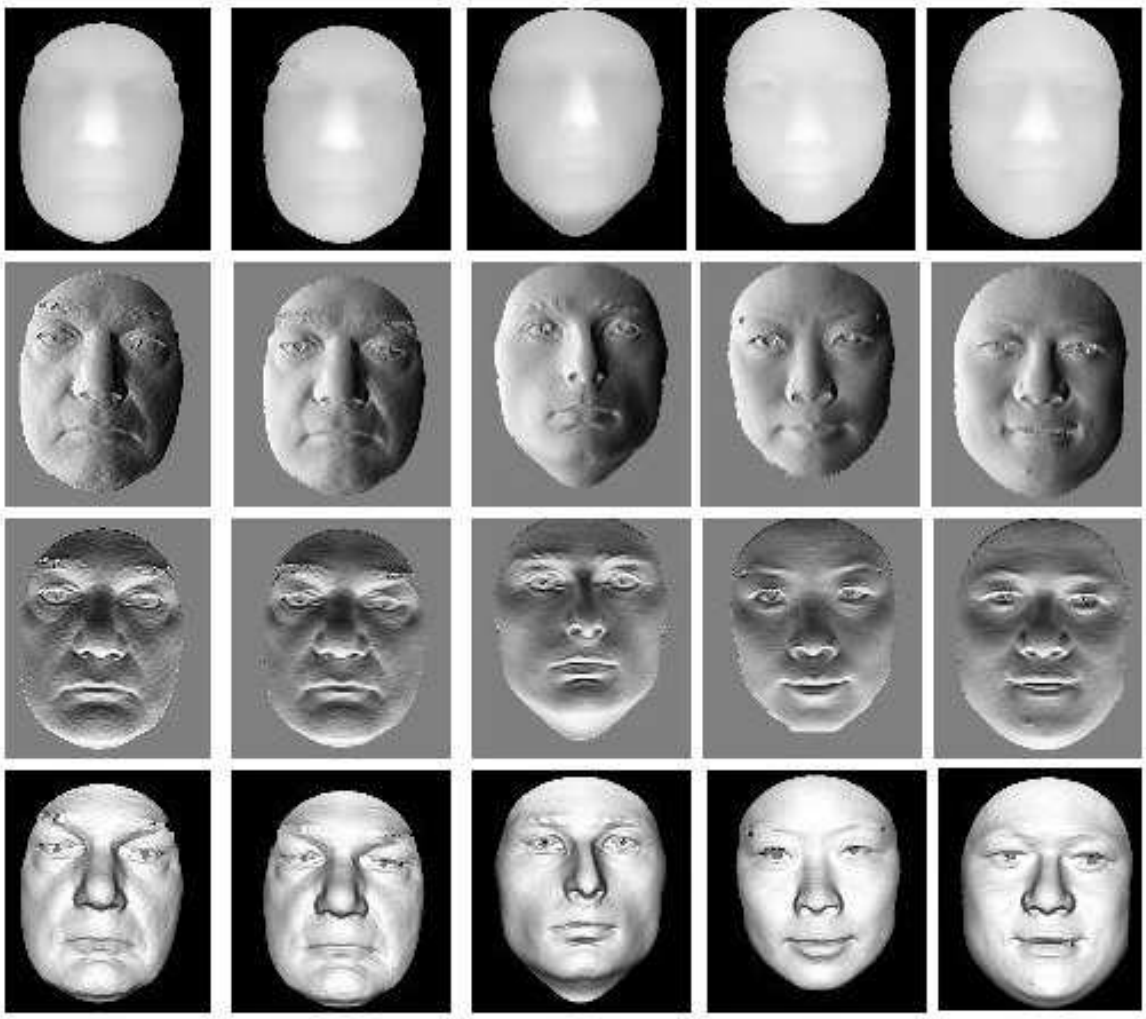

Figure 5.2 – Surface normal components for same subjects (first and second column) and different subjects (other columns) on FRGCv2, depth map,normal x, normal y, normal z in each row respectively

figure 5.3) as follow

$$f(I(p_i),I(q_{ij})) =$$

$$\begin{cases} 0, & \text{if } I(q_{ij}) - I(p_i) \leq threshold \\ 1, & \text{if } I(q_{ij}) - I(p_i) > threshold \end{cases}, j = 1,2,...,8 \tag{5.3}$$

LBP is a non-directional local pattern since it encodes all directions using first-order derivative pattern. However, LDP by extracting high-order derivative information provides more detailed distinct descriptor. Given normal components $N = [N_x, N_y, N_z]$, the first-order LNDP along four different directions including $0°, 45°, 90°$, and $135°$ is considered as $N'_\alpha(p_i) = [N'_{x\alpha}(p_i), N'_{y\alpha}(p_i), N'_{z\alpha}(p_i)]$ and calculated as follows

$$N'_{0°}(p_i) = N(p_i) - N(q_{i4}) \tag{5.4}$$

$$N'_{45°}(p_i) = N(p_i) - N(q_{i3}) \tag{5.5}$$

$$N'_{90°}(p_i) = N(p_i) - N(q_{i2}) \tag{5.6}$$

$$N'_{135°}(p_i) = N(p_i) - N(q_{i1}) \tag{5.7}$$

The second-order directional LNDP is defined as

$$\begin{aligned} LNDP^2_\alpha(p_i) = &\{f(N'_\alpha(p_i), N'_\alpha(q_{i1})), f(N'_\alpha(p_i), N'_\alpha(q_{i2})), ..., \\ &f(N'_\alpha(p_i), N'_\alpha(q_{i8}))\} \end{aligned} \tag{5.8}$$

where f is a binary coding function which is defined as follows

$$f(N'_\alpha(p_i), N'_\alpha(q_{ij})) = \begin{cases} 0, & \text{if } N'_\alpha(q_{ij}).N'_\alpha(p_i) > 0 \\ 1, & \text{if } N'_\alpha(q_{ij}).N'_\alpha(p_i) \leq 0 \end{cases}, j = 1, ..., 8 \tag{5.9}$$

Second-order LNDP is computed by concatenating four directions

$$LNDP^2(p_i) = \{LNDP^2_\alpha | \alpha = 0°, 45°, 90°, 135°\} \tag{5.10}$$

To calculate third-order local normal derivative pattern, first the second-order pattern using equation 5.8 is calculated and denoted as $N''(p_i)$ along four different directions

$$\begin{aligned} LNDP^3_\alpha(p_i) =& \{f(N''_\alpha(p_i), N''_\alpha(q_{i1})), f(N''_\alpha(p_i), N''_\alpha(q_{i2})), ..., \\ & f(N''_\alpha(p_i), N''_\alpha(q_{i8}))\} \end{aligned} \tag{5.11}$$

By applying concatenation of four directions we have

$$LNDP^3(p_i) = \{LNDP^3_\alpha | \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ\} \tag{5.12}$$

The general formula to calculate n^{th}-order normal derivative pattern is calculated by a binary string to describe gradient variations in a local area of $(n-1)^{th}$-order normal derivative pattern

$$\begin{aligned} LNDP^n_\alpha(p_i) =& \{f(N^{n-1}_\alpha(p_i), N^{n-1}_\alpha(q_{i1})), f(N^{n-1}_\alpha(p_i), \\ & N^{n-1}_\alpha(q_{i2})), ..., f(N^{n-1}_\alpha(p_i), N^{n-1}_\alpha(q_{i8}))\} \end{aligned} \tag{5.13}$$

where f is defined as

$$\begin{aligned} & f(N^{(n-1)}_\alpha(p_i), N^{(n-1)}_\alpha(q_{ij})) = \\ & \begin{cases} 0, & \text{if } N^{(n-1)}_\alpha(q_{ij}).N^{(n-1)}_\alpha(p_i) > 0 \\ 1, & \text{if } N^{(n-1)}_\alpha(q_{ij}).N^{(n-1)}_\alpha(p_i) \leq 0 \end{cases}, j = 1, ..., 8 \end{aligned} \tag{5.14}$$

and concatenating of four directions results in

$$LNDP^n(p_i) = \{LNDP^n_\alpha | \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ\} \tag{5.15}$$

The above equation defines each pixel of normal maps with a 32-bit binary encoding pattern. Figure 5.4 represents 32 templates to calculate binary functions of local derivative

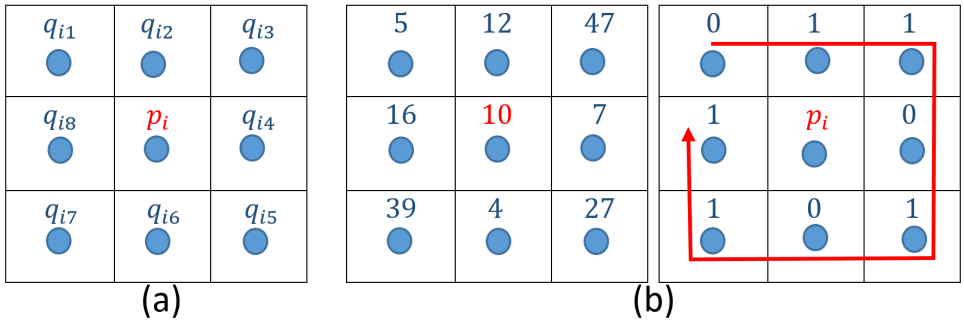

Figure 5.3 – (a) Eight neighbors around p_i, (b) LBP micro-pattern example

pattern on normal maps.

The figure shows from left to right to calculate
$f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i1})), f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i5})),$
$f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i2})), f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i6})),$
$f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i3})), f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i7})),$
$f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i4})), f(N_\alpha^{(n-1)}(p_i), N_\alpha^{(n-1)}(q_{i8})),$

and α is $0°, 45°, 90°, and 135°$ for first, second, third, and forth row respectively, $N = [N_x, N_y, N_z]$.

Since LDP encodes the different distinct spatial relationships in a local neighborhood of each point, it contains more spatial details compared to LBP to extract distinctive features. Spatial histogram *HLNDP* is applied to model the distributions of high-order local derivative pattern. Each normal images in x, y, and z direction is divided into L local patches and the histogram is extracted from each patch. The final descriptor is created using concatenation of all histograms extracted from each local patch (see figure 5.5).

$$HLNDP(l,\alpha) = \{HLNDP_\alpha(R_l) | l = 1, ..., L; \\ \alpha = 0°, 45°, 90°, 135°\} \tag{5.16}$$

Inspired by LBP, LNDP can be computed using different local neighborhood size in different scales. Multiscale LBP first has been proposed in [42] for texture classification and then used for 3D face recognition [64]. Around each central point p_i, a set of sampling points

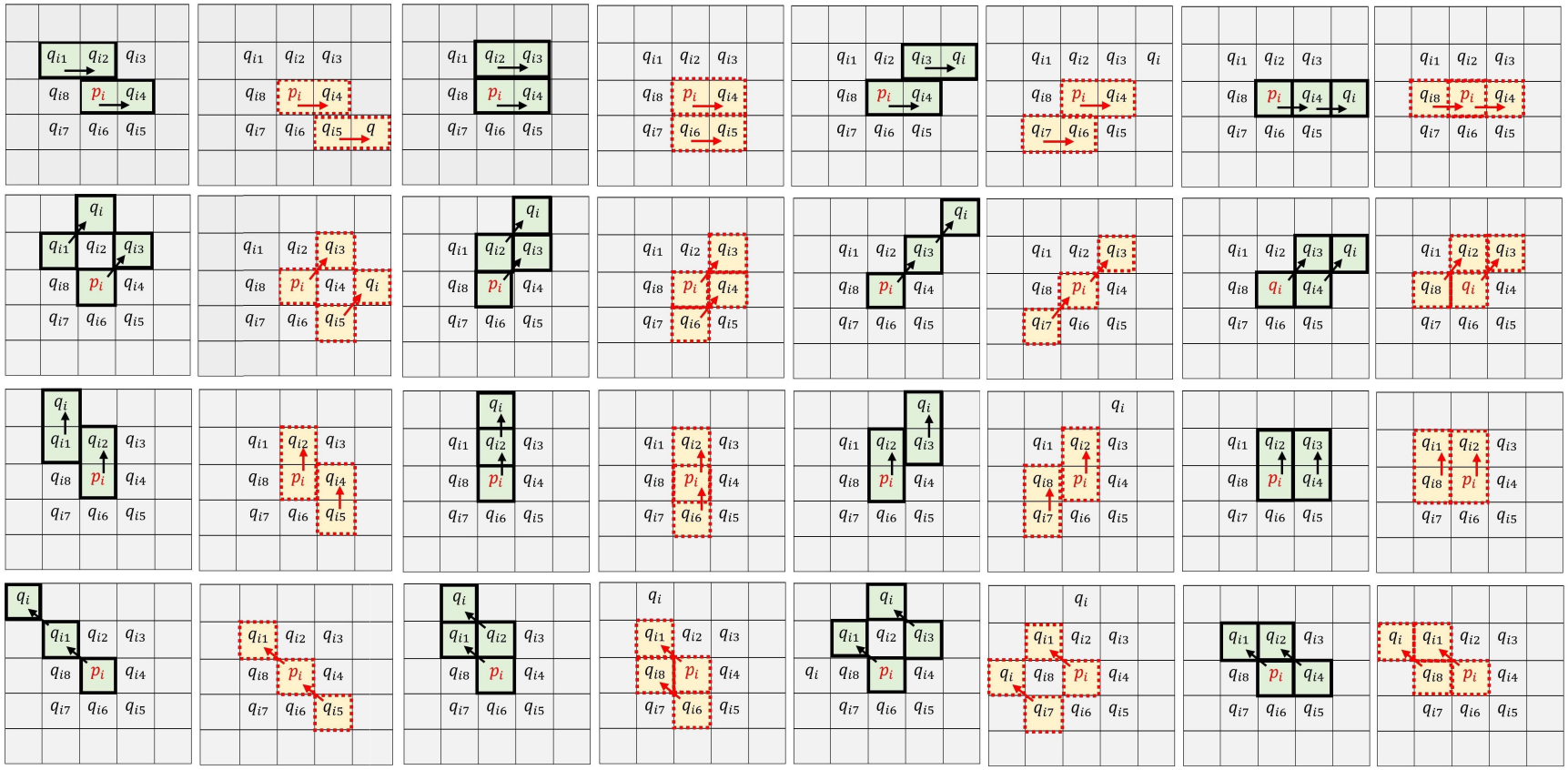

Figure 5.4 – 32 LNDP templates

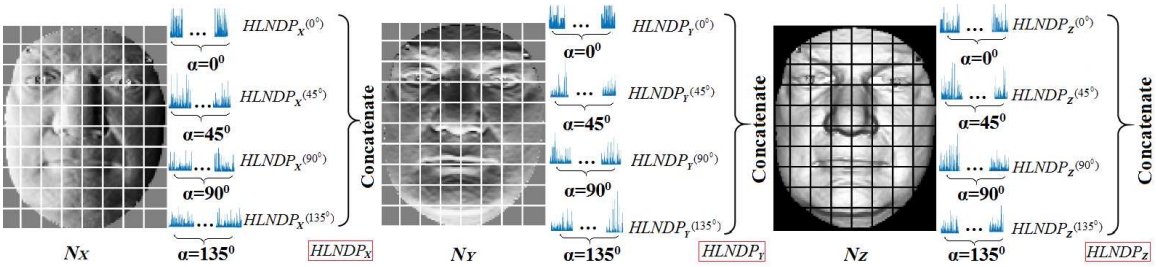

Figure 5.5 – Histogram of *LNDP* for x, y, and z normal maps

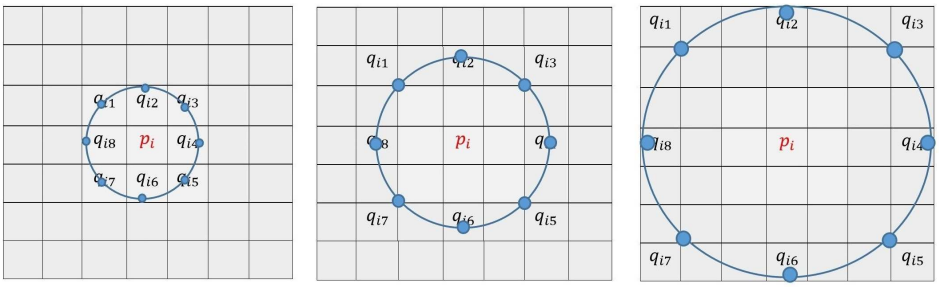

Figure 5.6 – Example of different scales, from left to right: $R = 1$, $R = 2$, and $R = 3$ for $U = 8$

are considered with different numbers U and radius R. As figure 5.6 illustrates multiscale LNDP is created by considering different values for R. We use 8 number of the neighbor points as U to compute LNDP.

The general form of the proposed high-order derivative descriptor, n^{th}-order MsLNDP is computed using the following equation

$$\begin{aligned} MsLNDP^{n}_{\alpha}(p_i) = \{ & f(N^{n-1}_{x,\alpha,R}(p_i), N^{n-1}_{x,\alpha,R}(q_{ij})), \\ & f(N^{n-1}_{y,\alpha,R}(p_i), N^{n-1}_{y,\alpha,R}(q_{ij})), f(N^{n-1}_{z,\alpha,R}(p_i), N^{n-1}_{z,\alpha,R}(q_{ij})) \\ & , j = 1, ..., U\} \end{aligned} \tag{5.17}$$

where U and R represents the different values for the neighboring points and radius to generate multiscale descriptor.

5.2.3 Dimension reduction

Histogram of the proposed descriptor on different patches for different angles results in a feature vector with high dimension for each sample. We apply the ELM-based multilayer architecture autoencoder [147] for dimension reduction that provides efficient generalization performance. Autoencoder is an artificial neural network that consists of an encoder and a decoder network. Transforming of the input data with high dimension into the feature space with lower dimension is done in encoder part. In the autoencoder applied in this work, the current weights of the encoding layer are replaced with the previous decoding layer to keep much correlation with the input data. Therefore, more distinct significant features can be simplified in this approach.

For given descriptors $d = [d_1, ..., d_M]$, a basic autoencoder objective is minimizing the reconstruction error J between inputs d and reconstructed outputs $\widehat{d}$, $J = \sum \|d_i - \widehat{d_i}\|, i = 1, ..., M$. Encoding and decoding process in a basic auto-encoder is defined as follows

$$\begin{aligned} H_f &= G(w_f, b_f, d) \\ \widehat{d} &= G(w_n, b_n, H_f) \end{aligned} \tag{5.18}$$

where w_f and w_n are encoding and decoding layer weights, b_f and b_n are defined as bias term, and G describes a general hidden neuron.

In a multilayer network, feature data which is the output of the hidden layer is computed as

$$H^i{}_f = G(w^i{}_f, b^i{}_f, H^{i-1}{}_f) \tag{5.19}$$

where $H^i{}_f$ and $H^{i-1}{}_f$ are the output data of the ith and $(i-1)$th layer respectively. Using the architecture in [147] with an invertible activation function G and maximum loop number LO the decoding layer parameters w_n, b_n are calculated based on the ELM equations and used to update the encoding layer parameters w_f, b_f and the feature data $H^i{}_f$. The final feature data $H^i{}_f$ is employed in next sections for our classifier input. The summary of the

Algorithm 5.1 The proposed MsLNDP descriptor

Input: 3D face data P
  for points in P **do**
    Calculate normal maps (N_x, N_y, and N_z)
  end for
  for different scales R and number of neighbors U **do**
    for each N **do**
      Divide into L patches
    end for
    for $alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ **do**
      for each patch **do**
        for each pixel in patch of N **do**
          Apply equation 5.17
        end for
        Histogram construction
      end for
      Concatenate the histogram for different patches
    end for
    Concatenate the histogram for different α
    return $HLNDP_x^n$, $HLNDP_y^n$, $HLNDP_z^n$
  end for
  ELM-based autoencoder for dimension reduction
  return $MsLNDP^n$ descriptor

proposed descriptor is presented in algorithm 5.1.

5.3 Proposed weighted hybrid classifier

In this section, we propose a hybrid classification method called weighted extreme sparse classifier, WESC, for 3D face recognition. The motivation of applying ELM and sparse representation methods is fast learning and ability to handle noisy and imperfect data (such as faces under occlusion and large pose variations). We believe that combination of ELM and SRC can improve recognition performance. Before presenting the hybrid classifier, in the following sub-sections, we briefly describe the concepts related to ELM and sparse representation.

5.3.1 Extreme Learning Machine

ELM is known as one of the state-of-the-art multiclass classification methods that works originally for single hidden layer feedforward networks (SLFNs). The hidden layer parameters (weights and biases) need not be tuned [134]. The following equation represents the objective function of ELM

$$min_{\beta} \|H(X)\beta - T\|_2^2 + \frac{1}{\lambda}\|\beta\|_2^2 \tag{5.20}$$

where $X = [x_1, x_2, ..., x_S]$ is a set of training samples, $H \in \mathbb{R}^{S \times P}$ denotes hidden layer output matrix with P nodes in the hidden layer, β is the output weight vector with length P, T denotes the class labels vector of length S, and λ is the regularization parameter. To solve the equation 5.20, the following solution is applied [148]

when $P > S$

$$\hat{\beta} = H^{\dagger}T = H^T(I\lambda + HH^T)^{-1}T \tag{5.21}$$

when $P < S$

$$\hat{\beta} = H^{\dagger}T = (I\lambda + H^TH)^{-1}H^TT \tag{5.22}$$

where $H^{\dagger}$ is the pseudo-inverse of H, H^T denotes the transpose of the H, and I is the identity matrix.

5.3.2 Sparse Representation

Sparse representation of a query sample y is estimated using dictionary atoms a_i that are extracted features from training data x_i.

$$y \approx a_1x_1 + a_2x_2 + a_3x_3 + ... + a_Nx_N \tag{5.23}$$

For N_i training samples for ith subject $[x_{i,1}, x_{i,2}, ..., x_{i,N_i}] \in \mathbb{R}^{M \times N_i}$, any test sample $y_i \in \mathbb{R}^M$ is

$$y_i \approx a_{i,1}x_{i,1} + a_{i,2}x_{i,2} + a_{i,3}x_{i,3} + ... + a_{i,N_i}x_{i,N_i} \tag{5.24}$$

where $a_{i,j} \in \mathbb{R}, j = 1, 2, ..., N_i$. Since based on the most common experimental protocols, there is only one training sample for each subject in the gallery for 3D face recognition system, the equation 5.23 is modified as

$$y_i = a_{i,1}x_{i,1} + \varepsilon \tag{5.25}$$

where $\varepsilon \in \mathbb{R}^M$ is an error term by different challenges. In this way, for N 3D faces in gallery (one sample per subject), the dictionary is defined as $D = [x_1, x_2, ..., x_N] \in \mathbb{R}^{M \times N}$ and any probe $y = Dc + \varepsilon$, where c is the coefficient vector approximated via the following optimization

$$\hat{c} = argmin_c \|y - Dc\|_2^2 \tag{5.26}$$

using a characteristic function δ_i to select the coefficient related to the ith gallery sample, the reconstruction residual is calculated

$$r_i(y) = \|y - D\delta_i(\hat{c})\|_2^2, i = 1, 2, ..., N \tag{5.27}$$

To find the probe label the minimum residual is applied

$$label(y) = argminr_i(y) \tag{5.28}$$

5.3.3 Weighted Extreme Sparse Classifier

In this section, a new weighted hybrid classifier is proposed to take advantage of the fast model training speed of ELM classifier and sparse representation capability to handle noisy images. Inspired by a combination of ELM and SRC for 2D face recognition in [149, 150],

a hybrid framework is proposed for 3D face scans. A discriminating criteria, the absolute difference between the first two largest elements of the ELM vector, is considered to make the hybrid classifier. If the difference is larger than a predefined threshold the classification is completed, otherwise, the query scan is reclassified by an adaptive sub-dictionary selection for SRC. Sub-dictionary is defined to classify the query sample and diminish computational cost effectively compared to applying the entire dictionary. We record the indexes of k largest elements in the ELM output. The sub-dictionary is constructed by picking up the atoms related to training samples with the same labels. The sub-dictionary is defined as $D_y^* = [D_{m(1)}, D_{m(2)}, ..., D_{m(k)}]$, where $m(i) \in 1,2,...,m$ denotes the indexes of the k largest entries. Applying sub-dictionary instead of computing the sparse coefficient over all training samples is as follows

$$\hat{c} = argmin_c \|y - D_y^* c\|_2^2 + \tau \|c\|_1 \tag{5.29}$$

where τ is the regularization parameter for SRC [149].

Inspired by weighted SRC to handle occlusion for 2D face [125] and expression handling for 3D face [15] a weighted framework for hybrid classifier is proposed to overcome various challenges. We divide each face into L different patches with learned weight w'. Feature vector is $x_i = [x_{i1}, x_{i2}, ..., x_{il}], l = 1,2,...,L$. ELM parameters are defined as $w_i = [w_{i1}, w_{i2}, ..., w_{il}]$, $b_i = [b_{i1}, b_{i2}, ..., b_{il}]$. The sub-dictionary is described as $D_y^* = [D_1^*, D_2^*, ..., D_l^*]$ and $D_{yl}^* = [x_{1,l}^*, x_{2,l}^*, ..., x_{N,l}^*]$ and any probe y can be written as $[y_1, ..., y_L]$.

Weight learning

To handle degradation conditions for face recognition the weight of each patch is learned to create a weighted hybrid classifier. Local patch weight learning has been applied in several 2D face recognition works [151, 125, 152]. These works show different regions of the face result in various contributions for the face recognition performance. Consequently, we apply different weights related to different patches in the hybrid classifier.

Algorithm 5.2 Proposed Weighted Extreme Sparse Classifier (WESC)

Input: A training database D with m classes, desired output $T_{N\times m}$, a query image y, regularized ELM parameter λ, regularized SRC parameter τ, threshold σ

Output: y class label

1: Hidden node parameters generation randomly (w_{il}, b_{il}), $i = 1, 2, ..., P$

2: Calculate

$$\sum_{l=1}^{L} w_l' H(w_{1l}, ..., w_{Pl}, x_{1l}, ..., x_{Nl}, b_{1l}, ..., b_{Pl})$$

3: Calculate weight matrix $\hat{\beta} = H^{\dagger}T$

4: ELM output O calculation for a probe sample y

$$O = \sum_{l=1}^{L} w_l' H(w_{1l}, ..., w_{Pl}, y_l, b_{1l}, ..., b_{Pl})\hat{\beta}$$

5: **if** $O_{first} - O_{second} > \sigma$ **then**

6: $label(y) = argmax(O)$

7: **else**

8: Find the indexes of k largest elements in O

9: Apply the weighted sub-dictionary D_{yl}^*

10: $\hat{c} = argmin_c \sum_{l=1}^{L} w_l' \|y_l - D_{yl}^* c\|_2^2 + \tau\|c\|_1$

11: **for** $i \in \{m(1), ..., m(k)\}$ **do**

12: Calculate $r_i(y) = \sum_{l=1}^{L} w_l' \|y_l - D_{yl}^* \delta(\hat{c})\|_2^2$

13: **end for**

14: $label(y) = argminr_i(y)$

15: **end if**

To learn weights for all following experiments, different datasets are applied. In the experiments, four different 3D face databases are used including FRGCv2 [16], Bosphorus [2], Bu-3DFE [49], and 3D-TEC [55]. Since the Bosphorus dataset has the highest variations in the expression, pose, and occlusion, it is used for weight learning in all experiments in which FRGCv2, BU-3DFE, and 3D-TEC are the test data. While to evaluate the performance on Bosphorus, BU-3DFE is employed for weight learning. By applying ESC classifier for different patch sizes on the database weights can be learned. From figure 5.7, it is obvious that the rigid areas such as the regions around nose and forehead have the highest weight and those ones that are sensitive under expression such as the regions

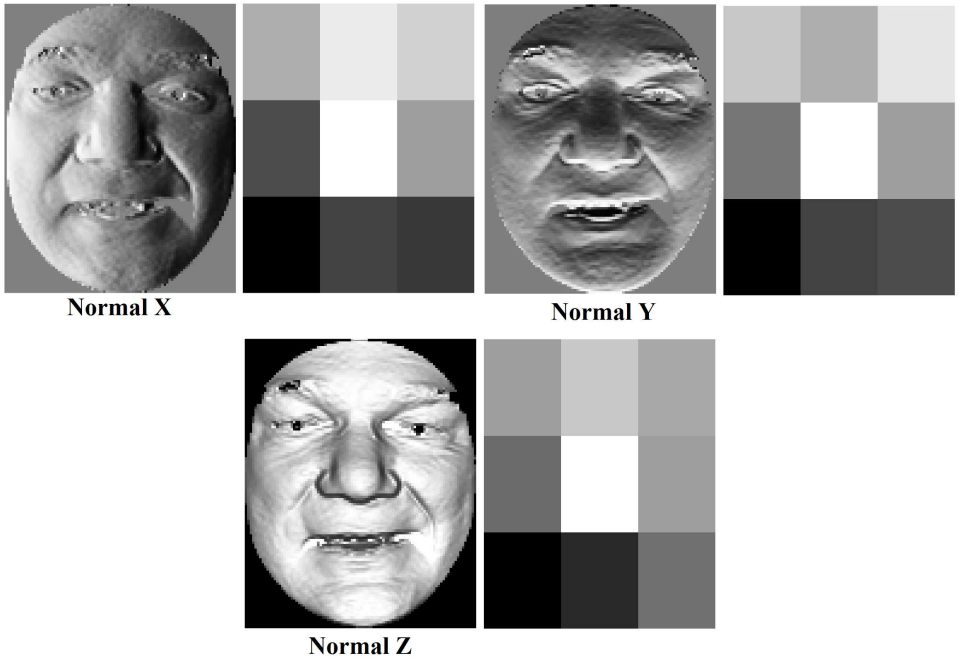

Figure 5.7 – Patch weights for three normal maps (x, y, and z direction) from FRGC database and local patch size equal to 40×32

around mouth have the smallest weight. The proposed hybrid classifier is summarized in algorithm 5.2.

5.4 Experiments and results

In this section the pre-processing step and evaluation of the proposed descriptor and classifier on four databases have been presented for 3D face recognition.

The most famous 3D face database, FRGC, contains two sets textured 3D face scans: v1 and v2 with 943 scans of 273 subjects and 4007 scans of 466 subjects respectively captured with the Minolta Vivid 900 scanner. The FRGCv1 scans are acquired with neutral expression and 640×480 resolution. The second one samples are captured under a limited range of facial expression (such as happiness and surprise) and controlled light and pose. Bosphorus database is made up of 4666 textured 3D scans of 105 subjects in an uncontrolled environment under different facial expression (neutral, happy, surprise, fear,

Table 5.1 – Specification of 3D databases used in this work

Name	Subjects/Scans	Scanner	Pose	Expression	Occlusion
FRGCv2[16]	466/4007	Laser	$\pm 15^{\circ}$	Mild to extreme	No
Bosphorus[2]	105/4666	Sterio	Yes	7 types	4 types
Bu-3DFE[49]	100/2500	Sterio	Frontal	6 types(4 levels)	No
3D-TEC[55]	214/428	Laser	Frontal	2 types	No

sadness, anger, and disgust), action units, poses, and occlusion. 3D scans are captured using Inspeck Mega Capturor II scanner and each range image has 1600×1200 resolution. BU-3DFE database contains 2500 3D facial scans of 100 subjects under 6 different expression including happiness, disgust, fear, anger, surprise, and sadness with four intensity levels and one neutral expression. The scans are acquired using 3D Imaging System (3DMD) and saved as a polygonal mesh with a resolution from 20000 to 35000 polygons. 3D-TEC database consists of scans from 214 subjects including 106 pairs of identical twins and a set of triplets with a neutral and smiling scan per subject. The specification and details of the four 3D databases applied in this work have been provided in table 5.1.

5.4.1 Pre-processing

There are some factors including noise which comes from the sensor and pose variation that can impact the rendered images. To diminish these factors and convert 3D models to high-quality 2D maps, the pre-processing is done on 3D scans using the pre-processing tool [128]. The median filter is applied to remove spike and noises. Hole filling is done by fitting square surface. We employ a curvature-based method to detect nose tip to crop region of interest (ROI). For the 3D-TEC dataset manually annotated nose tips positions are used. For pose correction, all the facial models are aligned together using a rigid-ICP algorithm [133]. We resized each pre-processed image into 120×96.

5.4.2 Performance of proposed descriptor

To evaluate the effectiveness of the proposed local derivative descriptor for 3D face recognition, a set of experiments on the largest 3D face database, FRGCv2, has been performed. The experiments are conducted in the MATLAB R2016a on a computer with an Intel Core i7, 3.60 GHz CPU with 8 GB RAM. Histogram intersection (HI) is used for recognition task to show the effectiveness of the proposed descriptor. The experiment is according to the protocol in [16] which the first scan of each subject is considered as a gallery and the remaining scans are used to make a probe set (neutral versus all).

First, to select the best dimension for extracted features we conduct the experiment to compare recognition results versus the number of features for depth, normal x, normal y, and normal z maps in figure 5.8. According to this figure, the best recognition rate is obtained for feature vectors with 600 dimensions for all descriptors. In other next experiments, we apply this value for feature dimension. In our experiments, we have set G as a sine function, $LO = 2$, and the number of layers $i = 4$.

Figure 5.9 demonstrates the recognition rate for various orders of the local derivative pattern on different maps. We employ local patches with 12×12 size to extract local pattern. According to the results, the recognition accuracy is effectively improved by increasing the order of local pattern from first-order to the second and third orders that means high order local pattern can extract more detailed distinct information from face data. However, by further increasing the order to forth-order, fifth-order, and sixth-order the accuracy drops that shows further detailed information contained in high-order local pattern converts face scan into noisy data and deteriorates the recognition rate.

Based on the above results, we employ third-order LDP in the next experiments. The effectiveness of different scales of LDP on different maps is evaluated and shown in figure 5.10. In all experiments, we use 8 neighbors around the central point to compute local derivative descriptor. In addition, we need to change the size of local patches for different scale size in this test. We tried six different radiuses to calculate LDP on normal images.

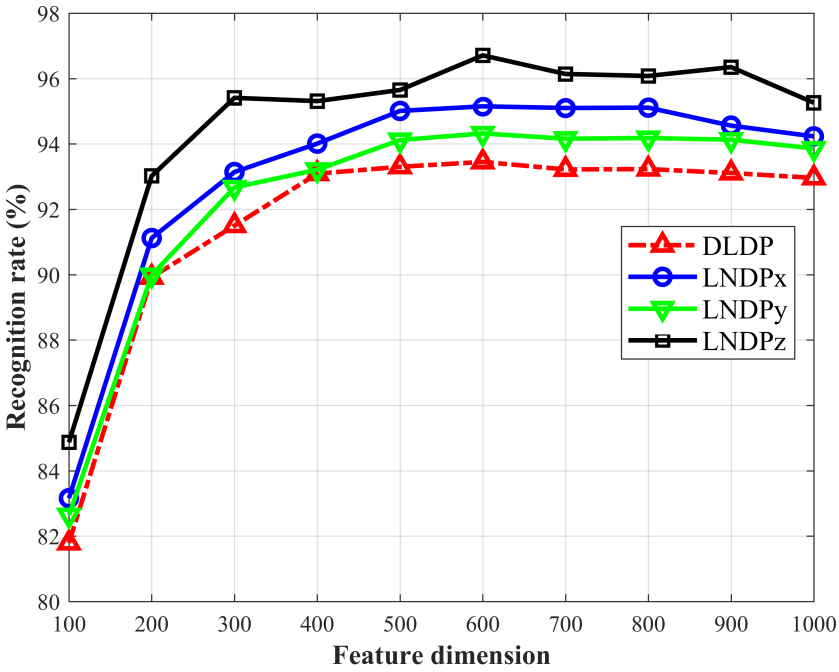

Figure 5.8 – Recognition accuracy versus feature dimension on FRGCv2 for depth (DLDP), normal x (LNDPx), normal y (LNDPy), and normal z (LNDPz)

For the two smallest scales $R = 1,2$ the local patches with the finest size, 12×12, for $R = 3,4$, patch size 20×16, and for $R = 5,6$, patch size 40×32 are considered. As the figure shows changing the scale to calculate LDP on depth and normal maps affect recognition performance. The scale 3 is the best one for all maps. The fusion of multiple encoding scales can enhance the recognition accuracy (A in figure 5.10). Based on the experiments we found that fusion of top three encoding scales based on the recognition rate provides the highest accuracy (T includes scales 2, 3, and 5 in figure 5.10). Score-level fusion with sum rule is applied to calculate the similarity of the multiscale descriptor. From figures 5.8-5.10 it is obvious that the LDP extraction on all three normal images outperforms the depth map. Using the above parameters, feature dimension: 600, order: 3, and multiscale scheme of scales: 2, 3, and 5, the fusion of three directions x, y, and z of the proposed pattern is considered as a final descriptor.

Finally, the comparison of the final proposed descriptor to other LBP-based methods

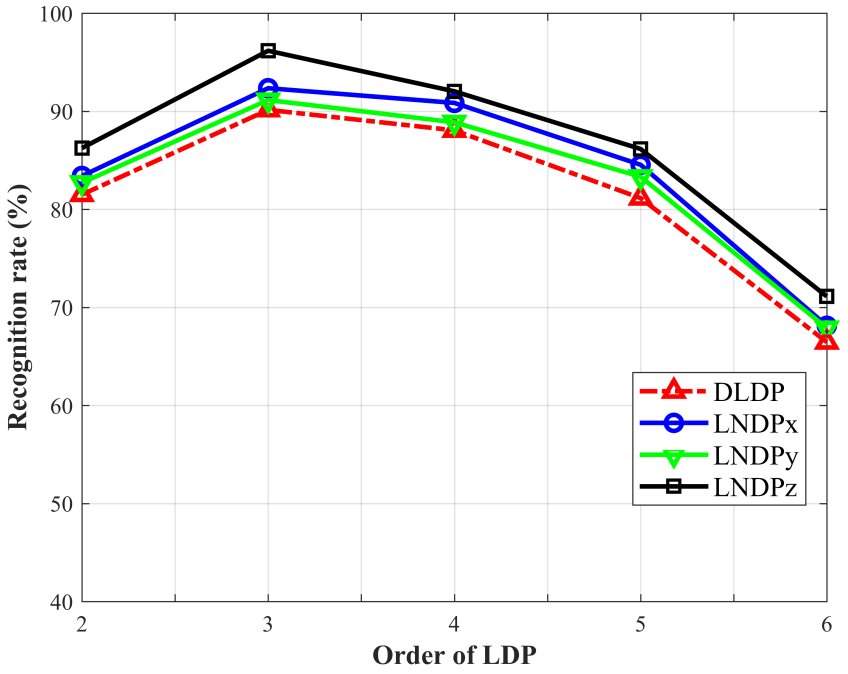

Figure 5.9 – Effectiveness of different orders of LDP on FRGCv2

including shape map and multiscale local binary pattern, SI+MS-LBP, [64], multiscale extended LBP, Ms-eLBP, and hybrid matching [81], the self-adaptive voting LBP, V-LBP [98], multiscale and multicomponent local normal patterns, MsMc-LNP, in combination with weighted SRC[15], multiscale depth local derivative pattern, MsDLDP, using SRC [143] and local normal derivative pattern, LNDPxyz, with HI [145] has been summarized in table 5.2. The neutral vs. all experiment is performed based on the same protocol on FRGCv2. As the results show our proposed enhanced descriptor is comparable to the state-of-the-art as well as our recent work [143] by applying ELM-based feature selection method and selective multiscale scheme. The high R1RR in [143] is for excluding non-rigid parts of the facial samples. While in this work, we have applied our proposed descriptor on whole faces.

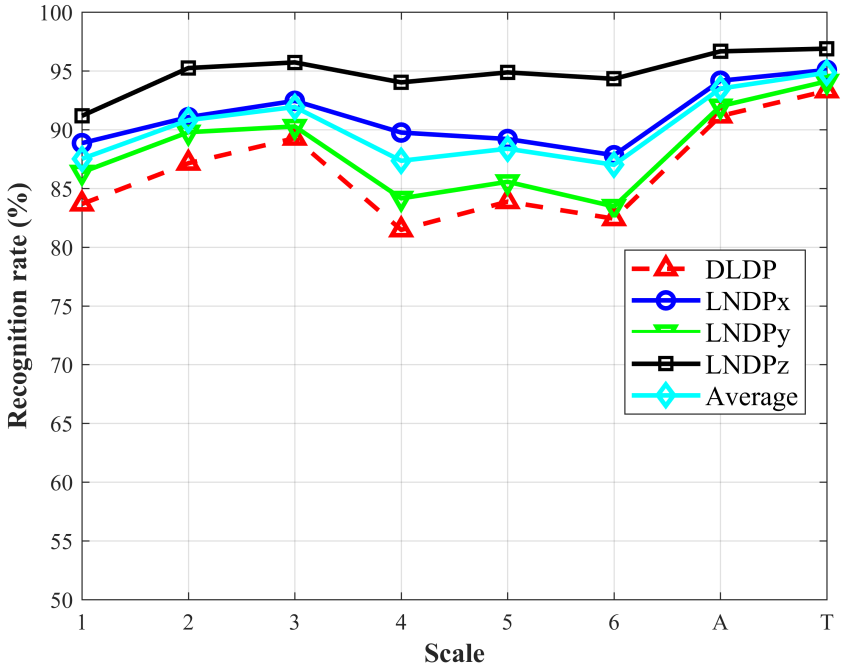

Figure 5.10 – Effectiveness of different scales of LDP on the recognition rate on FRGCv2, A: all scales are fused, T: scales 2, 3, and 5 (top three RR) are fused

5.4.3 Performance of proposed classifier

In this part to show the effectiveness of the proposed classifier, experimental results are conducted using multiscale depth local derivative pattern descriptor, MsDLDP.

To set the threshold σ and number of k for the proposed ESC, we calculated the R1RR and matching time per image on FRGCv2. As figure 5.11 depicts the recognition rate is improved by increasing σ and the number of k. However, according to the results in table 5.3 which report testing time per image for the different number of k, increasing the number of k causes computational cost. Moreover, it is obvious that a larger threshold results in higher recognition rate but more samples have been assigned to the time-consuming sparse classifier and computational complexity increases. Therefore, the best parameter setting is a trade-off between accuracy and time. Accordingly, we set the threshold value equal to 0.4 and $k = 200$. The regularized parameters for ELM and SRC is equal to 2 and 0.1 on

Table 5.2 – Comparison of the proposed descriptor with other LBP-based methods on FRGCv2

Method	R1RR
SI+Ms-LBP2010[64]	96.10%
Ms-eLBP2012[81]	97.60%
V-LBP2013[98]	94.90%
MsMc-LNP+WSRC2014[15]	96.30%
MsDLDP+SRC2017[143]	98.30%
Proposed descriptor+HI	**98.20%**

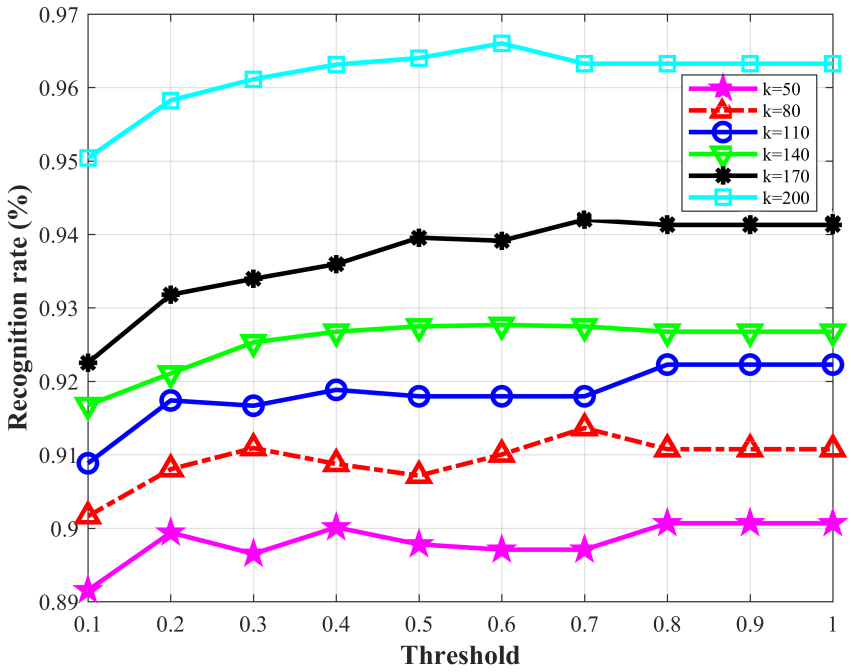

Figure 5.11 – R1RR of proposed ESC for different thresholds σ and different number k of largest entries

FRGCv2. By repeating the experiments we found that we can set the number of largest entries nearly equal to half of the number of gallery samples.

We also studied the affection of two various popular activation functions, the sigmoid and hyperbolic tangent functions, for the different number of hidden nodes to evaluate ELM and proposed ESC performance. Based on the results in figure 5.12, we choose a sigmoid function for the classifiers. In addition, the figure illustrates the effectiveness of the ESC

Table 5.3 – Testing time per image for different number k of largest entries

k	50	80	110	140	170	200
Testing time (sec)	0.0035	0.0052	0.0089	0.0134	0.0165	0.0227

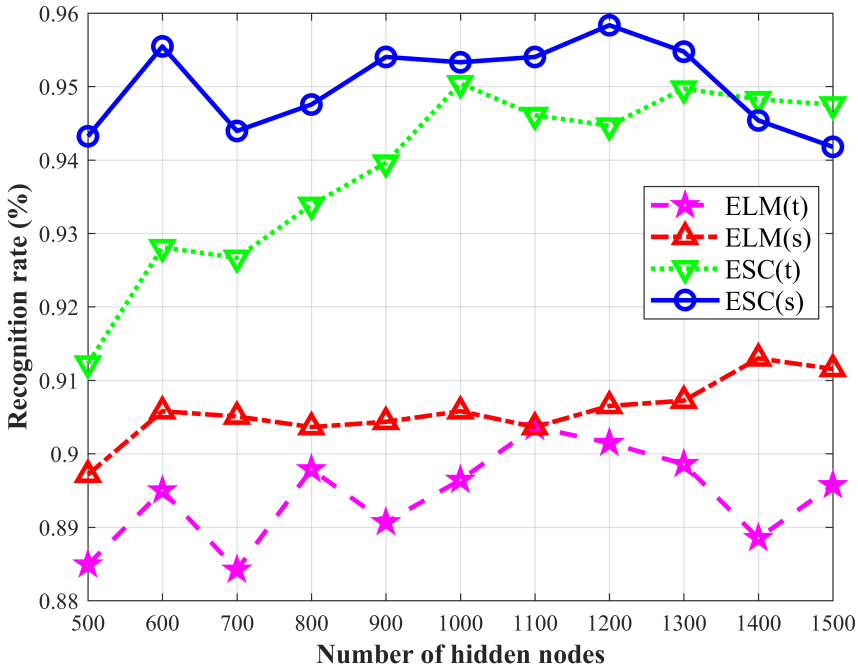

Figure 5.12 – R1RR using two activation functions

compared to ELM and the best result is obtained for 1200 hidden nodes.

The performance of the Weighted ESC and Non-weighted ESC is compared with other classifiers including ELM method, and SRC on FRGCv2 database using the MsDLDP descriptor. The results in table 5.4 show the effectiveness of the proposed method in terms of accuracy and speed. The proposed WESC classifier provides higher recognition rate compared to ELM and SRC classifiers. Not only combination of ELM and SRC improves computational complexity of the sparse representation classifier but also causes to increase the accuracy of the recognition system. Weighted scheme of the proposed ESC can handle facial challenges efficiently and has the highest classification accuracy among other methods.

Table 5.4 – Comparison of the proposed WESC algorithm with other classifiers on FRGCv2

Algorithm	R1RR	Training time (sec)	Testing time (sec)
ELM	93.89%	4.12	0.0063
SRC	95.21%	-	0.2450
ESC	95.83%	3.36	0.0227
WESC	97.65%	3.85	0.0290

5.4.4 Performance on different databases

According to the experiments in the previous sections, we continue the experiments on different 3D face databases using the fusion of three normal descriptors in the multiscale scheme, proposed MsLNDP, by employing proposed weighted ESC to overcome facial challenges. Table 5.5 compares the R1RR for different experiments according to the protocols in [16] on FRGCv2 with the state-of-the-art. The results depict the high quality of the proposed algorithm in all experiments. Compared with [15] and [95], that apply normal components as the basis of their feature space, there is 1.9% and 0.45% improvement in R1RR when the neutral samples are used as the probes. For non-neutral samples as the probe, the improvement of the proposed method is 0.1% and 4.4% relative to [95], and [15] respectively. Our method outperforms [95], and [15] with 1.4% and 3% improvement in R1RR when all samples made the probe.

Another criteria to evaluate the 3D face recognition approach is computational complexity. Table 5.6 reports matching time for the proposed method along with recognition rate and the comparison with state-of-the-art methods on FRGCv2 database for neutral vs. all experiment. As the table depicts our proposed method runs faster than other methods with higher recognition rate. Unlike other methods that compare a probe face with every gallery samples, our SRC-based method compares the probe with all gallery samples at the same time. In addition, feature dimension reduction and applying sub-dictionary instead of whole dictionary for sparse classifier could improve our method's processing time.

To more evaluate the performance of the proposed algorithm we continue the experi-

Table 5.5 – Comparison of 3D face recognition methods in term of R1RR on FRGCv2

Algorithm	neutral vs. neutral	neutral vs. non-neutral	neutral vs. all
Mian et al.2008[3]	99.40%	92.10%	96.10%
Huang et al.2011[127]	99.20%	95.10%	97.60%
Lei et al.2013[13]	-	95.6%	-
Drira et al.2013[10]	99.20%	96.80%	97.0%
Berretti et al.2013[90]	97.30%	92.80%	95.60%
Smeets et al.2013[4]	-	-	89.6%
Ocegueda et al.2013[113]	-	-	96.6%
Li et al.2014[15]	98.0%	94.20%	96.30%
Elaiwat et al.2015[7]	99.4%	94.1%	97.1%
Li et al.2015[70]	-	-	96.3%
Al-Osaimi2016[94]	99.1%	96.49%	97.78%
Soltanpour and Wu2016[66]	99.60%	96.0%	96.9%
Lei et al.2016[67]	99.6%	92.2%	96.3%
Emambakhsh and Evans2017[95]	98.45%	98.5%	97.9%
This work	**99.9%**	**98.6%**	**99.3%**

ments on Bosphorus dataset. For this experiment, the BU-3DFE dataset is used for weight learning. There are different experiments on this dataset due to expression, pose, and occlusion challenges. For test under expression, the gallery consists of 105 neutral samples belong to different subjects, and 2797 samples with neutral and non-neutral expression make up probes. Some researchers reported the results on 381 samples under pose, 3196 frontal samples under expression and occlusion challenges, and nearly 4561 samples under all challenges including expression, pose, and occlusion. We have evaluated our algorithm under expression and all challenges. A comparison with sate-of-the-art is provided in table 5.7. The table depicts remarkable performance of our method under facial challenges. The 2.93% drop in R1RR for all challenges (105/4561) compared to the experiment under facial expression (105/2797) shows the proposed method is sensitive under extreme pose and occlusion variations. However, our algorithm improves R1RR, 2.4%, and 2.35% compared to normal-based feature approaches [95, 15] respectively. Although the accuracy of the proposed algorithm on Bosphorus samples under pose variations and occlusion has

Table 5.6 – Comparison of 3D face recognition methods in term of matching time on FRGCv2

Algorithm	Matching time(s)	RR1
Huang et al.2011[127]	0.32	97.60%
Drira et al.2013[10]	1.27	97.0%
Berretti et al.2013[90]	0.2	95.60%
Li et al.2014[15]	0.5	96.30%
Elaiwat et al.2015[7]	0.36	97.1%
Lei et al.2016[67]	2.41	96.3%
Soltanpour and Wu2016[66]	0.35	96.9%
This work	**0.1**	**99.3%**

been reduced. However, our method achieves competitive performance compared to the state-of-the-art.

In tables 5.8 and 5.9, we evaluate our algorithm on BU-3DFE and 3D-TEC datasets and compare the results with others. On BU-3DFE which is challenging dataset for its samples with intense facial expression, the 100 neutral samples make up the gallery and the remaining samples under different types of expression create the probes. On 3D-TEC we apply the protocol in [55]. The database contains twins samples with neutral and smile expression. One person in each pair is labeled A and the other one B and four cases are applied for the experiment. In case 1, twin A with the smile and twin B with the smile, in case 2, neutral twin A and neutral twin B, in case 3, twin A with the smile and neutral twin B, and in case 4, neutral twin A and twin B with the smile make up the galley. In each case, the remaining samples make up the probes. According to the results, our method performs better or comparable with other works on these two databases. Comparison with the state-of-the-art on four different databases proves a successful generalization of our proposed method.

5.5 Conclusion

In this chapter, a novel classification method called WESC inspired by recent advances in extreme learning machine and sparse representation has been proposed. An adaptive

Table 5.7 – Comparison of 3D face recognition methods in term of R1RR on Bosphorus (”n: neutral”, ”e: expression”, ”o: occlusion”, ”p: pose”, ”all: e, o, p”).

Algorithm	Type(gallery/probe)	R1RR
Huang et al.2011[127]	(n/e,o)	97.0%
Li et al.2011[5]	(n/all)	94.1%
Drira et al.2013[10]	(n/o)	87%
Berretti et al.2013[69]	(n/all)	93.40%
Smeets et al.2013[4]	(n/e)	97.70%
	(n/all)	93.70%
Ocegueda et al.2013[113]	(n/all)	93.8%
Berretti et al.2014[6]	(n/all)	94.5%
Li et al.2014[15]	(n/e)	95.40%
Al-Osaimi2016[94]	(n/e)	92.41%
	(n/o)	84.78%
	(n/all)	90.28%
Soltanpour and Wu2016[66]	(n/e,o)	97.20%
	(n/all)	94.50%
Emambakhsh and Evans2017[95]	(n/e)	95.35%
This work	(n/e)	**97.75%**
	(n/all)	**94.82%**

Table 5.8 – Comparison of 3D face recognition methods in term of R1RR on BU-3DFE

Algorithm	R1RR
Mpiperis et al.2007[86]	84.4%
Berretti et al.2013[69]	87.5%
Li et al.2014[15]	92.21%
Berretti et al.2014[6]	88.2%
Werghi et al.2016[12]	93.42%
Lei et al.2016[67]	93.25%
Emambakhsh and Evans2017[95]	88.9%
Kim et al.2017[18]	95%
Li et al.2018[153]	95.25%
This work	**95.36**%

weighted sub-dictionary selection for SRC and regularized ELM was used to construct the classifier. In addition, a novel multiscale local derivative pattern has been proposed to further handle facial challenges by extracting distinct features. An ELM-based autoencoder has been employed to extract robust distinct features. Different experiments on four

Table 5.9 – Comparison of 3D face recognition methods in term of R1RR on 3D-TEC

Algorithm	3D-TEC			
	Case 1	Case 2	Case 3	Case 4
Faltemier et al.2008[108]	94.4%	93.5%	72.4%	72.9%
Huang et al.2010[64]	92.1%	93.0%	83.2%	83.2%
Li et al.2014[15]	95.8%	96.7%	95.3%	95.3%
Al-Osaimi2016[94]	95.79%	97.2%	87.38%	85.98%
Kim et al.2017[18]	94.8%	94.8%	81.3%	79.9%
This work	**96.3%**	**98.6%**	**97.1%**	**96.7%**

databases have been performed to evaluate the performance of the proposed algorithm under different scenarios. The experimental results reveal a reasonable generalization on different databases and performance improvement in term of accuracy and computational complexity. The proposed feature extraction and classifier can be applied to other 3D object recognition applications. In this work, we have employed the rigid-ICP method for face registration. Face registration algorithm to extract pose corrected face data is an interesting research area for the future.

Chapter 6

Conclusion and Future Directions

6.1 Conclusion

Recognition of 3D human faces is an interesting and active research area in the field of computer vision and pattern recognition. Despite the interest in this research field, handling deformations remained a challenging task due to non-rigid nature of the face. Extraction of geometric information of the 3D face data makes it more promising than 2D texture images since shape information is not sensitive to viewpoint and illumination variations.

There are three different categories of face recognition algorithms: holistic feature-based, local feature-based and hybrid methods. Recent advances in 3D face recognition with main focus on local feature-based methods were presented in chapter 2. Advantages and limitations of various 3D local feature-based methods were summarized for three different categories. According to the survey, local feature-based methods show more efficient results compared to holistic feature-based methods. Complete models are not required in local feature-based methods and consequently, occlusion can be handled. Some effective and robust 2D face descriptors such as SIFT and LBP can be applied on 3D maps to extract local descriptors which are robust under facial expression. Moreover, local descriptors can handle expression variations by excluding sensitive facial regions. In addition, local features can be detected on rigid patches or parts of the faces which are the least affected under expression. The holistic methods require accurate normalization for pose and scale. To normalize 3D data in holistic methods manual and automatic landmark detection is used which the manual landmarking is more accurate. However, it is time-consuming and makes the whole process semi-automatic. Moreover, pose normalization under noisy or

low-resolution 3D scans is a challenging task. Based on the literature, local feature-based methods are more suitable for matching, identification, and verification, since the main focus of local methods is on shape details. However, for similarity in search applications, the holistic features work better compared to local feature-based methods. In particular, the survey conducted in this dissertation shows no existing methods can handle all facial recognition challenges, which include expression, occlusion, missing data, and background clutter.

The third chapter was inspired by keypoint-based methods which are efficient under expression challenges and occlusion. A novel local descriptor was proposed to detect SIFT keypoints on shape maps in three different levels of the Gaussian pyramid to guarantee keypoints repeatability and provide more distinct features. SIFT keypoints were also detected on texture image to enhance the recognition system performance in a multimodal scheme. In addition, score level fusion was applied to calculate the final score using texture and shape modality's matching score. Experimental results depict that verification rate on FRGCv2 database has achieved 1% and 0.55% improvement compared to state-of-the-art for the most challenging experiments, all vs. all and ROCIII respectively. The improvement on Bosphorus database is equal to 4.8% for the verification rate and 0.4% for identification rate.

In the fourth chapter, 3D face recognition using LBP-based local surface methods was presented. Since LBP is an efficient local descriptor in 2D face recognition applications, a novel multiscale high-order local pattern called MsDLDP was proposed; which can handle facial expression by excluding non-rigid parts of the face and sparse representation-based classifier. The proposed descriptor contains more spatial information compared to LBP by encoding the various distinct spatial relationships in a local neighborhood. The multiscale strategy was proposed to enhance the effectiveness of the detected features. A comparison was performed on sparse representation and distance-based classifier results. The proposed algorithm using SRC could enhance recognition rate 9.84% on FRGCv2 database and 11.55% on Bosphorus database compared to Chi-square classifier. The comparison

with LBP-based methods illustrates 0.5% and 0.45% recognition rate improvement for FRGCv2 and Bosphorus databases respectively.

In the last chapter, the local derivative pattern on surface normals in the x, y, and z directions called MsLNDP using a weighted hybrid classifier was presented. Moreover, an ELM-based dimension reduction method was applied to extract distinct features. A learning-based framework was considered to calculate local patch weights to handle different facial challenges. A combination of SRC and ELM classifiers called weighted extreme sparse classifier was proposed by learning an ELM network and adopting a discriminant criteria to decide about the ELM output reliability. In the case of unreliable output, the features are fed into SRC to extract sub-dictionary and reduce computational burden. The proposed WESC could improve the recognition rate and testing time 2.44% and 0.216*sec* compared to SRC on FRGCv2 database respectively. The proposed algorithm including MsLNDP and WESC achieved 1.4% improvement for neutral vs. all experiment on FRGCv2 database. In term of computational complexity the proposed method could enhance the matching speed twice for neutral vs. all experiment on FRGCv2 database. On Bosphorus database, the proposed method could provide 0.32% improvement for the recognition rate for neutral vs. all experiment. On BU-3DFE database, the improvement by 0.11% and on 3D-TEC database, the average improvement by 1.4% have been achieved for the recognition rate.

6.2 Suggestions for Future Work

There are a number of areas which future research could explore, they are as follows:

- The local feature-based methods for 3D face recognition have been surveyed using the experiment results from other works. For further investigation, each local feature-based method can be implemented and evaluated on different databases having different challenges.

- The main focus of this dissertation was local feature extraction and classification

algorithms. The existing tools for pre-processing and pose correction were applied in this work. In the future, new algorithms for face pre-processing, registration and pose correction could be investigated. Further work could be conducted on landmarking algorithms for nose tip detection and segmentation of the region of interest. Denoising, hole filling, and spike removal methods can also be areas of future work.

- The proposed local pattern was applied on depth and surface normal maps. In the future, the proposed descriptor can be applied on pyramidal shape maps to investigate the generalization of the proposed local pattern.

- In this work, an ELM-based auto-encoder was used for dimension reduction. Investigating other methods to extract more distinct and robust features is another interesting and useful research direction.

- The last algorithm applied patch-based weight learning to handle facial challenges. In the future, different algorithms such as GA for weight learning could be used and the results compared.

- The presence of artifacts and incomplete facial data can create challenges in practical applications of 3D local feature-based methods. Moreover, 3D face data acquisition is computationally more expensive than 2D data capturing. Therefore, handling artifacts and 3D data acquisition need more attention to improve.

- Unavailability of a large-scale 3D face database which contains a combination of different challenges including extreme expression, pose variation, and occlusion is a major limitation in the 3D face recognition area. Therefore, a creation of such a 3D face database could be an important task to be done.

- It would be interesting to analyze the proposed methods' sensitivity using low-resolution 3D samples or the 3D samples approximated from 2D scans. The reason behind this is although 3D laser scanners have been decreasing the cost, many of the existing databases are in 2D modality.

References

[1] M. Bennamoun, Y. Guo, and F. Sohel, "Feature selection for 2D and 3D face recognition," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2015.

[2] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *European Workshop on Biometrics and Identity Management*. Springer, 2008, pp. 47–56.

[3] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 1–12, 2008.

[4] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "meshSIFT: Local surface features for 3D face recognition under expression variations and partial data," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158–169, 2013.

[5] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen, "Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities," in *IEEE International Conference on Image Processing*, 2011, pp. 3053–3056.

[6] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, "Selecting stable keypoints and local descriptors for person identification using 3D face scans," *The Visual Computer*, vol. 30, no. 11, pp. 1275–1292, 2014.

[7] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, "A curvelet-based approach for textured 3D face recognition," *Pattern Recognition*, vol. 48, no. 4, pp. 1235–1246, 2015.

[8] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 80–95, 2009.

[9] C. Samir, A. Srivastava, and M. Daoudi, "Three-dimensional face recognition using shapes of facial curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1858–1863, 2006.

[10] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.

[11] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," *Pattern Recognition*, vol. 47, no. 2, pp. 509–524, 2014.

[12] N. Werghi, C. Tortorici, S. Berretti, and A. Del Bimbo, "Boosting 3D LBP-based face recognition by fusing shape and texture descriptors on the mesh," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 964–979, 2016.

[13] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition*, vol. 46, no. 1, pp. 24–37, 2013.

[14] S. Berretti, M. Daoudi, P. Turaga, and A. Basu, "Representation, analysis and recognition of 3D humans: A survey," *ACM Transactions on Multimedia Computing, Communications and Applications*, 2018.

[15] H. Li, D. Huang, J.-M. Morvan, L. Chen, and Y. Wang, "Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," *Neurocomputing*, vol. 133, pp. 179–193, 2014.

[16] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 947–954.

[17] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, "Deep face recognition." in *British Machine Vision Conference (BMVC)*, vol. 1, no. 3, 2015, p. 6.

[18] D. Kim, M. Hernandez, J. Choi, and G. Medioni, "Deep 3d face identification," *arXiv preprint arXiv:1703.10714*, 2017.

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

[20] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognitiona review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005.

[21] C. Creusot, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3D meshes," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 146–179, 2013.

[22] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.

[23] S. Soltanpour, B. Boufama, and Q. J. Wu, "A survey of local feature methods for 3D face recognition," *Pattern Recognition*, vol. 72, pp. 391–406, 2017.

[24] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches to three-dimensional face recognition," in *IEEE International Conference on Pattern Recognition (ICPR)*, 2004, pp. 358–361.

[25] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A survey of 3D face recognition methods," in *International Conference on Audio-and Video-based Biometric Person Authentication*. Springer, 2005, pp. 891–899.

[26] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[27] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[28] D. Smeets, P. Claes, D. Vandermeulen, and J. G. Clement, "Objective 3D face recognition: Evolution, approaches and challenges," *Forensic Science International*, vol. 201, no. 1, pp. 125–132, 2010.

[29] H. Zhou, A. Mian, L. Wei, D. Creighton, M. Hossny, and S. Nahavandi, "Recent advances on singlemodal and multimodal face recognition: a survey," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 6, pp. 701–716, 2014.

[30] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, no. 3, p. 37, 2016.

[31] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens, "A comparative study of 3D face recognition under expression variations," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 5, pp. 710–727, 2012.

[32] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image and Vision Computing*, vol. 30, no. 10, pp. 683–697, 2012.

[33] C. A. Corneanu, M. Oliu, J. F. Cohn, and S. Escalera, "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.

[34] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 1391–1398.

[35] X. Lu and A. Jain, "Deformation modeling for robust 3D face matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1346–1357, 2008.

[36] Y. Wang, J. Liu, and X. Tang, "Robust 3D face recognition by local shape difference boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1858–1870, 2010.

[37] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3D face recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 914–925, 2013.

[38] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 381–397, 2013.

[39] S. Z. Gilani, A. Mian, and P. Eastwood, "Deep, dense and accurate 3D face correspondence for generating population specific deformable models," *Pattern Recognition*, vol. 69, pp. 238–250, 2017.

[40] N. Bayramoglu and A. A. Alatan, "Shape index SIFT: range image recognition using local features," in *IEEE International Conference on Pattern Recognition (ICPR)*, 2010, pp. 352–355.

[41] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[42] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[43] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *IEEE International Conference on Advanced video and signal based surveillance (AVSS)*, 2009, pp. 296–301.

[44] A. Mian and N. Pears, "3D face recognition," in *3D Imaging, Analysis and Applications*, 2012, pp. 311–366.

[45] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, "3D assisted face recognition: A survey of 3D imaging, modelling and recognition approachest," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2005, pp. 114–114.

[46] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *IEEE International Symposium on Signal Processing and its Applications*, 2003, pp. 201–204.

[47] A. B. Moreno and A. Sánchez, "Gavabdb: a 3D face database," in *Workshop on Biometrics on the Internet*, 2004, pp. 75–80.

[48] C. Conde, Á. Serrano, and E. Cabello, "Multimodal 2D, 2.5D & 3D face verification," in *IEEE International Conference on Image Processing*, 2006, pp. 2061–2064.

[49] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2006, pp. 211–216.

[50] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition using combinations of surface feature map subspace components," *Image and Vision Computing*, vol. 26, no. 3, pp. 382–396, 2008.

[51] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+ 3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, 2005.

[52] C. Xu, T. Tan, S. Li, Y. Wang, and C. Zhong, "Learning effective intrinsic features to boost 3D-based face recognition," in *European Conference on Computer Vision*. Springer, 2006, pp. 416–427.

[53] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using a multi-instance enrollment representation to improve 3D face recognition," in *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2007, pp. 1–6.

[54] F. B. Ter Haar, M. Daoudi, and R. C. Veltkamp, "Shape retrieval contest 2008: 3D face scans," in *Shape Modeling International*, 2008, pp. 225–226.

[55] V. Vijayan, K. Bowyer, and P. Flynn, "3D twins and expression challenge," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2011, pp. 2100–2105.

[56] R. C. Veltkamp, S. van Jole, H. Drira, B. B. Amor, M. Daoudi, H. Li, L. Chen, P. Claes, D. Smeets, J. Hermans *et al.*, "Shrec'11 track: 3D face models retrieval," in *3DOR*, 2011, pp. 89–95.

[57] A. Colombo, C. Cusano, and R. Schettini, "UMB-DB: A database of partially occluded 3D faces," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2011, pp. 2113–2119.

[58] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3D face recognition database," in *IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*, 2010, pp. 97–100.

[59] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," in *IEEE International Conference On Automatic Face & Gesture Recognition (FG)*, 2008, pp. 1–6.

[60] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, "A comprehensive performance evaluation of 3D local feature descriptors," *International Journal of Computer Vision*, vol. 116, no. 1, pp. 66–89, 2016.

[61] S. Berretti, A. del Bimbo, and P. Pala, "3D partial face matching using local shape descriptors," in *Proceedings of the Joint ACM Workshop on Human Gesture and Behavior Understanding*. ACM, 2011, pp. 65–71.

[62] M. Mayo and E. Zhang, "3D face recognition using multiview keypoint matching," in *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2009, pp. 290–295.

[63] T. Inan and U. Halici, "3D face recognition with local shape descriptors," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 577–587, 2012.

[64] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, and L. Chen, "3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching," in *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, 2010, pp. 1–7.

[65] S. Soltanpour, Q. J. Wu, and M. Anvaripour, "Multimodal 2D-3D face recognition using structural context and pyramidal shape index," in *IET International Conference on Imaging for Crime Prevention and Detection (ICDP)*, 2015, pp. 2–6.

[66] S. Soltanpour and Q. J. Wu, "Multimodal 2D-3D face recognition using local descriptors: pyramidal shape map and structural context," *IET Biometrics*, vol. 6, no. 1, pp. 27–35, 2016.

[67] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, and X. Zhou, "A two-phase weighted collaborative representation for 3D partial face recognition with single sample," *Pattern Recognition*, vol. 52, pp. 218–237, 2016.

[68] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens, and D. Vandermeulen, "Feature detection on 3D face surfaces for pose normalisation and recognition," in *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, 2010, pp. 1–6.

[69] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, "Matching 3D face scans using interest points and local histogram descriptors," *Computers & Graphics*, vol. 37, no. 5, pp. 509–525, 2013.

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

[71] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 31–43, 2006.

[72] P. Perakis, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection under large yaw and expression variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1552–1564, 2013.

[73] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recognition*, vol. 39, no. 3, pp. 444–455, 2006.

[74] C. Creusot, N. Pears, and J. Austin, "Automatic keypoint detection on 3D faces using a dictionary of local shapes," in *IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2011, pp. 204–211.

[75] M. L. Koudelka, M. W. Koch, and T. D. Russ, "A prescreener for 3D face recognition using radial symmetry and the hausdorff fraction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 168–168.

[76] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3D face recognition," *International Journal of Computer Vision*, vol. 90, no. 3, pp. 331–349, 2010.

[77] M. Song, D. Tao, S. Sun, C. Chen, and S. J. Maybank, "Robust 3D face landmark localization based on local coordinate coding," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5108–5122, 2014.

[78] F. M. Sukno, J. L. Waddington, and P. F. Whelan, "3D facial landmark localization with asymmetry patterns and shape regression from incomplete local features," *IEEE Transactions on Cybernetics*, vol. 45, no. 9, pp. 1717–1730, 2015.

[79] M. A. de Jong, A. Wollstein, C. Ruff, D. Dunaway, P. Hysi, T. Spector, F. Liu, W. Niessen, M. J. Koudstaal, M. Kayser *et al.*, "An automatic 3D facial landmarking algorithm using 2D Gabor Wavelets," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 580–588, 2016.

[80] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, 2014.

[81] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1551–1565, 2012.

[82] S. Berretti, A. Del Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2162–2177, 2010.

[83] S. Jahanbin, H. Choi, Y. Liu, and A. C. Bovik, "Three dimensional face recognition using iso-geodesic and iso-depth curves," in *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2008, pp. 1–6.

[84] I. Mpiperis, S. Malasiotis, and M. G. Strintzis, "3D face recognition by point signatures and iso-contours," in *International Conference on Signal Processing, Pattern Recognition, and Applications*. ACTA Press, 2007, pp. 328–332.

[85] L. Li, C. Xu, W. Tang, and C. Zhong, "3D face recognition by constructing deformation invariant image," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1596–1602, 2008.

[86] I. Mpiperis, S. Malassiotis, and M. G. Strintzis, "3D face recognition with the geodesic polar representation." *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3-2, pp. 537–547, 2007.

[87] S. Feng, H. Krim, and I. Kogan, "3D face recognition using euclidean integral invariants signature," in *IEEE Workshop on Statistical Signal Processing*, 2007, pp. 156–160.

[88] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava, "Pose and expression-invariant 3D face recognition using elastic radial curves," in *British Machine Vision Conference*, 2010, pp. 1–11.

[89] L. Ballihi, B. B. Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3D geometric features for efficient face recognition and gender classification," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1766–1779, 2012.

[90] S. Berretti, A. Del Bimbo, and P. Pala, "Sparse matching of salient facial curves for recognition of 3D faces with missing parts," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 374–389, 2013.

[91] X. Li and F. Da, "Efficient 3D face recognition handling facial expression and hair occlusion," *Image and Vision Computing*, vol. 30, no. 9, pp. 668–679, 2012.

[92] P. J. Besl, N. D. McKay *et al.*, "A method for registration of 3D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[93] L. Zhang, A. Razdan, G. Farin, J. Femiani, M. Bae, and C. Lockwood, "3D face authentication and recognition based on bilateral symmetry analysis," *The Visual Computer*, vol. 22, no. 1, pp. 43–55, 2006.

[94] F. R. Al-Osaimi, "A novel multi-purpose matching representation of local 3d surfaces: A rotationally invariant, efficient, and highly discriminative approach with an adjustable sensitivity," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 658–672, 2016.

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

[96] S. Z. Li, C. Zhao, M. Ao, and Z. Lei, "Learning to fuse 3D+ 2D based face recognition at both feature and decision levels," in *International Workshop on Analysis and Modeling of Faces and Gestures*. Springer, 2005, pp. 44–54.

[97] Y. Huang, Y. Wang, and T. Tan, "Combining statistics of geometrical and correlative features for 3D face recognition," in *BMVC*. Citeseer, 2006, pp. 879–888.

[98] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Processing*, vol. 93, no. 8, pp. 2190–2198, 2013.

[99] N. Werghi, S. Berretti, and A. Del Bimbo, "The mesh-LBP: a framework for extracting local binary patterns from discrete manifolds," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 220–235, 2015.

[100] C. Xu, Y. Wang, T. Tan, and L. Quan, "Automatic 3D face recognition combining global geometric features with local shape variation information," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 308–313.

[101] X. Li and H. Zhang, "Adapting geometric attributes for expression-invariant 3D face recognition," in *IEEE International Conference on Shape Modeling and Applications (SMI)*, 2007, pp. 21–32.

[102] X. Li, T. Jia, and H. Zhang, "Expression-insensitive 3D face recognition using sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2575–2582.

[103] H. Tabia, H. Laga, D. Picard, and P.-H. Gosselin, "Covariance descriptors for 3D shape matching and retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 4185–4192.

[104] W. Hariri, H. Tabia, N. Farah, A. Benouareth, and D. Declercq, "3D face recognition using covariance based descriptors," *Pattern Recognition Letters*, vol. 78, pp. 1–7, 2016.

[105] C. S. Chua, F. Han, and Y. K. Ho, "3d human face recognition using point signature," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 233–238.

[106] Y. Wang and C.-S. Chua, "Robust face recognition from 2D and 3D images using structural hausdorff distance," *Image and Vision Computing*, vol. 24, no. 2, pp. 176–185, 2006.

[107] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695–1700, 2006.

[108] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "A region ensemble for 3D face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 62–73, 2008.

[109] A. S. Mian, M. Bennamoun, and R. A. Owens, "Matching tensors for pose invariant automatic 3D face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 120–120.

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

[111] Y. Ming, "Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis," *Image and Vision Computing*, vol. 35, pp. 14–22, 2015.

[112] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, 2007.

[113] O. Ocegueda, T. Fang, S. K. Shah, and I. A. Kakadiaris, "3D face discriminant analysis using Gauss-Markov posterior marginals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 728–739, 2013.

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

[115] D. Huang, W. B. Soltana, M. Ardabilian, Y. Wang, and L. Chen, "Textured 3D face recognition using biological vision-based facial representation and optimized

weighted sum fusion," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011, pp. 1–8.

[116] G. Zhang and Y. Wang, "Robust 3D face recognition based on resolution invariant features," *Pattern Recognition Letters*, vol. 32, no. 7, pp. 1009–1019, 2011.

[117] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Spatially optimized data-level fusion of texture and shape for face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 859–872, 2012.

[118] W. Liu and Y. Yang, "Structural context for object categorization," in *Pacific-Rim Conference on Multimedia*. Springer, 2009, pp. 280–291.

[119] N. Alyuz, B. Gokberk, and L. Akarun, "Regional registration for expression resistant 3D face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 425–440, 2010.

[120] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[121] J. J. Koenderink, *Solid shape*. MIT press, 1990.

[122] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[123] M. Emambakhsh and A. Evans, "Nasal patches and curves for an expression-robust 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016*, 2016.

[124] S. Lv, F. Da, and X. Deng, "A 3d face recognition method using region-based extended local binary pattern," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3635–3639.

[125] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[126] Y. Ming and Q. Ruan, "Robust sparse bounding sphere for 3d face recognition," *Image and Vision Computing*, vol. 30, no. 8, pp. 524–534, 2012.

[127] D. Huang, K. Ouji, M. Ardabilian, Y. Wang, and L. Chen, "3D face recognition based on local shape patterns and sparse representation classifier," in *International Conference on Multimedia Modeling*. Springer, 2011, pp. 206–216.

[128] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6.

[129] Z. Zhang, "Iterative point matching for registration of free-form curves," Ph.D. dissertation, Inria, 1992.

[130] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533–544, 2010.

[131] C.-H. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," in *International Conference on Biometrics*. Springer, 2007, pp. 809–818.

[132] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in *European Conference on Computer Vision*. Springer, 2010, pp. 749–762.

[133] U. Castellani and A. Bartoli, "3D shape registration," in *3D Imaging, Analysis and Applications*. Springer, 2012, pp. 221–264.

[134] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2012.

[135] I. Marques and M. Graña, "Face recognition with lattice independent component analysis and extreme learning machines," *Soft Computing*, vol. 16, no. 9, pp. 1525–1537, 2012.

[136] K. Choi, K.-A. Toh, and H. Byun, "Incremental face recognition for large-scale social network services," *Pattern Recognition*, vol. 45, no. 8, pp. 2868–2883, 2012.

[137] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine," *Neural Processing Letters*, vol. 39, no. 1, pp. 25–43, 2014.

[138] A. Baradarani, Q. J. Wu, and M. Ahmadi, "An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering," *Pattern Recognition*, vol. 46, no. 1, pp. 57–72, 2013.

[139] S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Multi-layer group sparse coding for concurrent image classification and annotation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2809–2816.

[140] J. Yang, L. Zhang, Y. Xu, and J.-y. Yang, "Beyond sparsity: The role of l1-optimizer in pattern classification," *Pattern Recognition*, vol. 45, no. 3, pp. 1104–1118, 2012.

[141] H. Tang, Y. Sun, B. Yin, and Y. Ge, "3D face recognition based on sparse representation," *The Journal of Supercomputing*, vol. 58, no. 1, pp. 84–95, 2011.

[142] G. Pan, X. Zhang, Y. Wang, Z. Hu, X. Zheng, and Z. Wu, "Establishing point correspondence of 3D faces via sparse facial deformable model," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4170–4181, 2013.

[143] S. Soltanpour and Q. J. Wu, "Multiscale depth local derivative pattern for sparse representation based 3d face recognition," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 560–565.

[144] R. S. Llonch, E. Kokiopoulou, I. Tošić, and P. Frossard, "3D face recognition with sparse spherical representations," *Pattern Recognition*, vol. 43, no. 3, pp. 824–834, 2010.

[145] S. Soltanpour and Q. J. Wu, "High-order local normal derivative pattern (lndp) for 3d face recognition," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 2811–2815.

[146] K. Klasing, D. Althoff, D. Wollherr, and M. Buss, "Comparison of surface normal estimation methods for range sensing applications," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 3206–3211.

[147] Y. Yang, Q. J. Wu, and Y. Wang, "Autoencoder with invertible functions for dimension reduction and image reconstruction," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–15, 2016.

[148] G.-B. Huang, "What are extreme learning machines? filling the gap between frank rosenblatts dream and john von neumanns puzzle," *Cognitive Computation*, vol. 7, no. 3, pp. 263–278, 2015.

[149] J. Cao, K. Zhang, M. Luo, C. Yin, and X. Lai, "Extreme learning machine and adaptive sparse representation for image classification," *Neural Networks*, vol. 81, pp. 91–102, 2016.

[150] M. Luo and K. Zhang, "A hybrid approach combining extreme learning machine and sparse representation for image classification," *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 228–235, 2014.

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

[152] Z. Lei, S. Liao, M. Pietikainen, and S. Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 247–256, 2011.

[153] Y. Li, Y. Wang, J. Liu, and W. Hao, "Expression-insensitive 3D face recognition by the fusion of multiple subject-specific curves," *Neurocomputing*, vol. 275, pp. 1295–1307, 2018.

IEEE Permission to Reprint

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University of Windsor's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to

http://www.ieee.org/publications standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

Vita Auctoris

Sima Soltanpour was born in Iran, in 1982. She received her BSc degree in Electrical Engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran in 2005 and her MSc degree in Electrical Engineering from the Sahand University of Technology, Tabriz, Iran in 2010. She is currently pursuing her PhD degree in Electrical Engineering with the University of Windsor, Windsor, Ontario, Canada and is graduating in 2018. Pattern recognition, image processing, computer vision, machine learning, and biometrics are her main research interests.