Towards three-dimensional face recognition in the real

Huibin Li

ÉCOLE CENTRALE LYON

LIRIS

THESE

pour obtenir le grade de
DOCTEUR DE L'ECOLE CENTRALE DE LYON
Spécialité: Informatique

Towards Three-Dimensional Face Recognition in the Real

---

dans le cadre de l'Ecole Doctorale InfoMaths
présentée et soutenue publiquement par

HUIBIN LI

November 18, 2013

Directeur de thèse: Prof. Liming CHEN
Co-directeur de thèse: Prof. Jean-Marie MORVAN

Dr. Bruno LEVY	INRIA	Rapporteur
Prof. Pietro PALA	Université de Florence	Rapporteur
Prof. Alice CAPLIER	Institut Polytechnique de Grenoble	Président
Dr. Stéphane GENTRIC	Morpho Safran	Examinateur
Prof. Xianfeng DAVID GU	State University New York at Stony Brook	Examinateur
Prof. Liming CHEN	Ecole Centrale de Lyon	Directeur de thèse
Prof. Jean-Marie MORVAN	Université Lyon 1	Co-directeur de thèse

Acknowledgments

Completing a PhD thesis is like a marathon, and I would not have been able to finish this marathon without the support of countless people over the past four and a half years.

First and foremost, I would like to express my deeply-felt thanks to my advisor, Prof. Liming Chen, for his continuous support and encouragement, and also for his direction and countless hours spent with me on this thesis and my research. Working under the supervision of such a knowledgeable and understanding person has been a great pleasure for me. Meanwhile, I would like to express my deeply-felt thanks to my co-advisor, Prof. Jean-Marie Morvan, for his guidance of learning the normal cycle theory, and also for his attitude of doing valuable research. He is knowledgeable, smart, funny and easy going; excellent qualifications for my co-advisor. Moreover, I would also like to express my deeply-felt thanks to Prof. Xianfeng David Gu, for teaching me the interest topic of computational conformal mapping, and for helping me to finish the surface meshing work. He is knowledgeable, quite hardworking, and energetic researcher.

I would also like to express my great gratitude to each of my committee members: Dr. Bruno Levy, Prof. Pietro Pala, Prof. Alice Caplier, Dr. Stephane Gentric, and Prof. Xianfeng David Gu, for their time, interest and helpful comments. Meanwhile, I give my great thanks to prof. Su San, who is native English speaker, for his very nice help to check the English of the main part (chapter 3 to chapter 5) of my thesis.

I would like to acknowledge the members of the LIRIS lab at the Ecole Centrale de Lyon, for their friendship, nice help, thoughtful suggestions, and contributions to my research. They are Dr. Huanzhang Fu, Dr. Xi Zhao, Dr. Yan Liu, Dr. Di Huang, Dr. Chao Zhu, Dr. Pierre Lemaire, Dr. Wale Ben Soltana, Dr. Karima Ouji, Dr. Chu Duc Nguyen, Dr. Przemyslaw Szeptycki, Boyang Gao, Ningning Liu, Yu Zhang and Tao Xu. My special thanks are due to Dr. Di Huang for guiding me to be a 3D face recognition guy, for his helpful suggestions, luminous ideas, and for all those unforgettable moments that we worked together. I am also grateful to my friends at the Ecole Centrale de Lyon, Dr. Gang Niu and Dr. Zhugen Yang, for all those unforgettable times we spend together. I am also grateful to my friends: Dr. Wei Zeng, Dr. Xiang Zeng and Lei Fan, for all their nice help when I was visiting the State University New York at Stony Brook, and also for the valuable discussions with Dr. Xiang Zeng, pleasant cooperation with Dr. Wei Zeng.

Finally, I would like to express my deepest gratitude and love to my family. My father, Chenjun Li, always to be the greatest person in my heart. He give me the opportunity to receive the higher education and also change my life. My mother, Hailu Gao, always give me her deepest love, and support me for the countless years of my academic career. My brother, Shoubin Li, he always spend much more times than me to take care of my mother during the past eight years. I am particularly indebted to my wife, Anbo Wu, for her deeply love, continuous support and encouragement. She has made my life happy, exciting and fun, and without her I would have had many more stressful and worrisome moments.

I would like to acknowledge the Chinese Government and the China Scholarship Council, without their financial support, it would not have been possible for me to study in France. My research in this PhD thesis is supported by the French research agency, Agence Nationale de Recherche (ANR) and Natural Science Foundation of China (NSFC), under the 3D Face Analyzer Project, grant ANR 2010 INTB 0301 01 and grant NSFC 61061130560.

Contents

Abstract xxi

Résumé xxiii

1 Introduction 1
- 1.1 Context and Motivation . . . 1
  - 1.1.1 What is Biometrics? . . . 1
  - 1.1.2 Why Face Recognition? . . . 4
  - 1.1.3 Face Recognition: A Brief Review in 2D . . . 6
  - 1.1.4 Face Recognition: Challenges in 2D . . . 9
  - 1.1.5 Face Recognition: Opportunities in 3D . . . 13
- 1.2 3D Face Recognition . . . 13
  - 1.2.1 3D Face Acquisition Techniques . . . 13
  - 1.2.2 Basic Concepts and Terminology . . . 15
  - 1.2.3 3D Face Databases . . . 17
  - 1.2.4 Main Challenges and Objectives . . . 18
- 1.3 Thesis Contributions . . . 21
  - 1.3.1 Methodologies . . . 21
  - 1.3.2 Main Contributions . . . 22
- 1.4 Thesis Organization . . . 27

2 Literature Review: 3D Face Recognition 29
- 2.1 Expression Deformation Modeling based Approaches . . . 29
  - 2.1.1 Thin-Plate-Spline (TPS) based Deformation Modeling . . . 30
  - 2.1.2 Annotated Deformable Model based Deformation Modeling . . . 32
  - 2.1.3 PCA Subspace based Deformation Modeling . . . 38
  - 2.1.4 3D Morphable Model based Deformation Modeling . . . 40
  - 2.1.5 Bilinear Model based Deformation Modeling . . . 41
- 2.2 Surface-distance based Approaches . . . 43
  - 2.2.1 Iso-metric Surface based Approaches . . . 43
  - 2.2.2 Iso-level Curves based Approaches . . . 45
  - 2.2.3 Iso-geodesic Strips based Approaches . . . 48
- 2.3 Local Region/Feature based Approaches . . . 49
  - 2.3.1 Multiple Region Matching based Approaches . . . 50
  - 2.3.2 Local Geometry-Texture Descriptor based Approaches . . . 57
  - 2.3.3 Local Geometry-Shape Descriptor based Approaches . . . 60
- 2.4 Discussion . . . 64
- 2.5 Summary . . . 65

3 3D Face Recognition in the Presence of Various Expressions 67
3.1 Introduction . . . 67
3.1.1 Related work . . . 68
3.1.2 Motivations and the proposed approach . . . 69
3.2 Overview of the Proposed Approach . . . 72
3.3 Local Normal Patterns (LNP) based Facial Descriptor . . . 73
3.3.1 Facial Normal Estimation . . . 73
3.3.2 Facial Normal Encoding . . . 75
3.3.3 Facial Normal Representation . . . 76
3.4 Weighted Sparse Representation-based Classifier . . . 77
3.4.1 3D Face Subspace and Sparse Representation-based Classifier . . . 77
3.4.2 Weighted Sparse Representation-based Classifier . . . 78
3.5 Experiments . . . 80
3.5.1 Databases and Preprocessing . . . 80
3.5.2 Experimental Settings . . . 82
3.5.3 Experimental Results . . . 84
3.6 Conclusion and Future Work . . . 94

4 3D Face Recognition in the Presence of Expression, Pose and Occlusion Variations 97
4.1 Introduction . . . 98
4.1.1 Overview of the Proposed Method . . . 100
4.1.2 Main Contributions of the Proposed Method . . . 101
4.2 Related Work . . . 103
4.2.1 3D Facial Representations . . . 103
4.2.2 Facial Features . . . 104
4.2.3 Matching Scheme . . . 105
4.3 Estimating Multiple Order Facial Surface Differential Quantities . . . 106
4.3.1 Facial Surface Gradient Estimator . . . 106
4.3.2 Facial Surface Curvature Estimator . . . 107
4.3.3 Facial Curvature Derivation Estimator . . . 108
4.4 3D Facial Salient Vertices Detection and their Local Shape Description . . . 109
4.4.1 3D Facial Salient Vertices Detection: κ_{max} and κ_{min} based Detectors 109
4.4.2 Pose-invariant Local Shape Description: Histograms of Multi-order Surface Differential Quantities . . . 110
4.5 Facial Surface Matching: a SRC-based Fine Grained Matcher vs. a Coarse Grained Matcher . . . 112
4.5.1 The Coarse Grained Matcher (CGM) . . . 113
4.5.2 SRC-based Fine Grained Matcher (FGM) . . . 113
4.6 Experimental Results . . . 116
4.6.1 Databases . . . 116
4.6.2 Experimental Setups . . . 119
4.6.3 Analysis of Salient Vertices Located by κ_{max} and κ_{min} Detectors . . . 121
4.6.4 Face Recognition on the Bosphorus Database . . . 124
4.6.5 Face Recognition on the FRGC v2.0 Database . . . 133
4.6.6 Discussion . . . 136

4.7 Conclusion . . . 141

5 3D Data Quality Improvement: Surface Meshing with Curvature Convergence 143
5.1 Introduction . . . 143
5.1.1 Existing Theoretical Results . . . 144
5.1.2 Our Theoretical Results . . . 145
5.2 Previous Works . . . 147
5.2.1 Surface Meshing/Remeshing . . . 147
5.2.2 Conformal Surface Parameterization . . . 148
5.3 Statement of the Main Theorem . . . 148
5.3.1 Curvature Measures . . . 148
5.3.2 Main Results . . . 149
5.4 Theoretical Proofs . . . 150
5.4.1 Surface Uniformization . . . 150
5.4.2 Delaunay Refinement . . . 151
5.4.3 Normal Cycle Theory . . . 153
5.4.4 Estimation . . . 156
5.5 Computational Algorithm . . . 164
5.5.1 Algorithm Pipeline . . . 166
5.6 Experimental Results . . . 167
5.6.1 Triangulation Quality . . . 167
5.6.2 Curvature Measure Comparisons . . . 168
5.7 Conclusion . . . 170

6 Conclusions and Future Work 171
6.1 Contributions . . . 171
6.1.1 3D Face Recognition in the Presence of Various Expressions . . . 171
6.1.2 3D Face Recognition in the Presence of Expression, Pose and Occlusion Variations . . . 172
6.1.3 3D Data Quality Improvement: Surface Meshing with Curvature Convergence . . . 172
6.2 Perspectives for Future Work . . . 173
6.2.1 3D Face Recognition: Encoding Surface Normal or Coordinate? . . . 173
6.2.2 3D Partial Face Recognition with Data Quality Guarantee . . . 173
6.2.3 3D Video-based Face Recognition . . . 174

A Smooth Surfaces 175
A.1 Local Differential Geometry . . . 175
A.1.1 Regular parameterized Surface Patch . . . 175
A.1.2 First Fundamental Form and Area . . . 176
A.1.3 Normal, Second Fundamental Form and Curvatures . . . 177
A.1.4 Surface Theory with Differential Forms . . . 180
A.2 Riemann Surface: Global Differential Geometry . . . 181
A.2.1 Riemann Surface and conformal mapping . . . 181
A.2.2 Ricci Flow . . . 186
A.2.3 Degree of Maps . . . 187

B Discrete Surfaces **189**
B.1 Local Differential Geometry . . . 189
B.1.1 Simplicial Complex and Triangular Mesh . . . 189
B.1.2 Discrete Normal and Curvatures . . . 190
B.2 Global Differential Geometry . . . 192
B.2.1 Discrete Metric . . . 192
B.2.2 Discrete Euclidean Ricci Flow . . . 193

C Surface Discretization and Approximation **195**
C.1 Surface Discretization . . . 195
C.1.1 Delaunay Triangulation . . . 195
C.1.2 Ruppert's Delaunay Refinement Algorithm . . . 195
C.2 Surface Approximations . . . 197
C.2.1 Two Negative Examples . . . 197
C.2.2 Unified Curvature Measures . . . 202
C.2.3 Curvature Measure Approximations . . . 204

D 3D Facial Expression Recognition Based on Histograms of Surface Differential Quantities **207**
D.1 Introduction . . . 207
D.2 Estimating Differential Quantities on Triangular Meshes . . . 208
D.2.1 Estimating Curvature by Normal Cycle Theory based Method . . . 208
D.2.2 Estimating Curvature by Local Cubic Fitting based Method . . . 209
D.2.3 Mesh Gradient Estimation . . . 211
D.3 Local Shape Descriptors . . . 211
D.3.1 Landmarks Selection . . . 211
D.3.2 Local Coordinate System and Orientation Assignment . . . 211
D.3.3 Feature Vectors Computed as Histograms of Surface Differential Quantities . . . 212
D.4 Experimental Results . . . 213
D.5 Conclusion and Future Work . . . 216

E 3D Facial Expression Recognition via Multiple Kernel Learning of Multi-Scale Local Normal Patterns **217**
E.1 Introduction . . . 217
E.2 Multi-scale local normal patterns . . . 218
E.3 Multi-class multiple kernel learning . . . 219
E.4 Experimental results . . . 221
E.5 Conclusion . . . 222

F An Automatic 3D Expression Recognition Framework based on Sparse Representation of Conformal Images **225**
F.1 Introduction . . . 225
F.1.1 Background . . . 225
F.1.2 Related Work . . . 226
F.1.3 Framework Overview . . . 227

F.2 Surface conformal mapping . . . 228
F.2.1 Conformal Mapping and Surface Conformal Representation . . . 228
F.2.2 Surface Ricci Flow . . . 229
F.3 Expression feature extraction: normalized conformal images . . . 229
F.3.1 Conformal Mapping Computation . . . 230
F.3.2 Conformal Mapping Normalization . . . 230
F.3.3 Conformal Images Computation . . . 231
F.4 Sparse representation of conformal image for expression classification . . . 231
F.4.1 Conformal Images based Expression Dictionary . . . 231
F.4.2 Sparse Representation of Conformal Images . . . 232
F.4.3 Expression Classification by Class-dependent Reconstruction Error 232
F.5 Experimental results . . . 233
F.5.1 Database and Preprocessing . . . 233
F.5.2 Experimental Settings . . . 233
F.5.3 Experimental Results . . . 233
F.6 Conclusion . . . 239

G A Group of Facial Normal Descriptors for Recognizing 3D Identical Twins 241
G.1 Introduction . . . 241
G.2 Framework Overview . . . 243
G.3 Facial Normal Estimation . . . 243
G.4 Facial Normal Encoding and Representation . . . 244
G.4.1 Normal Local Binary Patterns . . . 245
G.4.2 Normal Gabor Filters . . . 245
G.4.3 Normal Local Gabor Binary Patterns . . . 246
G.5 Sparse Representation Classifier . . . 247
G.6 Experimental and Algorithmic Settings . . . 248
G.6.1 Datasets and Preprocessing . . . 248
G.6.2 Experimental Settings . . . 249
G.6.3 Algorithmic Settings . . . 250
G.6.4 Experimental Results . . . 250
G.7 Conclude and Perspective . . . 252

H Publications 255

Bibliography 257

List of Tables

1.1 *A brief introduction of some public 3D databases.* 18

2.1 *Three types of facial curves based facial surface representations.* 46
2.2 *Riemannian analysis of curve spaces for facial curves and surfaces comparisons.* 47

3.1 *Experimental settings of FRGC v1.0, BU-3DFE, Bosphorus, and FRGC v2.0 databases.* 83
3.2 *Experimental setting of 3D-TEC database: "A Smile, B Neutral" means that the set contains all images with Twin A smiling and Twin B neutral [Vijayan et al. 2011b].* 84
3.3 *Comparison of rank-one scores: original normal, LSP and LNP on the whole FRGC v2.0 database.* 85
3.4 *Comparison of rank-one scores: Chi-Square vs. SRC on the whole FRGC v2.0 database.* 85
3.5 *The effectiveness of multi-scale and multi-component local normal patterns on the whole FRGC v2.0 database. The encoding scales, namely $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$, are computed over regular grid 10x8 with patch size 12x12, grid 6x6 with patch size 20x16, and grid 3x3 with patch size of 40x32, respectively. The range image size is normalized to 120x96 in this work. Fusion is carried out at score level through simple sum rule.* 87
3.6 *Comparison of rank-one score improvements on the FRGC v 2.0 database: patch weights are learned using FRGC v1.0, BU-3DFE, and Bosphorus respectively.* 89
3.7 *Comparison of rank-one score improvements on the BU-3DFE database: patch weights are learned from Bosphorus.* 89
3.8 *Comparison of rank-one score improvements on the Bosphorus database: patch weights are learned from BU-3DFE.* 89
3.9 *Comparing the degradations of rank-one scores influenced by facial expression changes on the FRGC v 2.0 database (Subset I: neutral probes; Subset II: non-neutral probes).* 90
3.10 *The degradations of rank-one scores of the proposed approach as the intensity of facial expressions increases in the BU-3DFE database.* 90
3.11 *Comparison of the rank-one scores on 3D-TEC by using different training sets.* 91
3.12 *Comparison of the Rank-1 scores on the FRGC v2.0, Bosphrous, BU-3DFE and 3D-TEC databases.* 92
3.13 *Computation time for preprocessing and/or feature extraction, face matching and identification of a single probe using a gallery of 466 subjects of the FRGCv2 database.* 94

4.1 *The distribution of 3D face scans over the various subsets in the Bosphorus database. In the subset of facial expressions, 105 neutral scans, one for each 105 subjects, are used as gallery set whereas the remaining ones as probe sets. The subset YR45° of size 210 (YR90°,* resp.*) contains both the subset of 105 scans at Yaw +45° (Yaw +90°,* resp.*) and those at Yaw −45° (Yaw −90°,* resp.*).* 119
4.2 *Parameter Values* 120
4.3 *Selection of the parameter μ for the coarse grained matcher (CGM): test on the whole Bosphorus database (105/4561), k_{max} + k_{min} salient vertex detectors.* 120
4.4 *Selection of the sparsity parameter L for the SRC-absed fine grained matcher (FGM): test on the whole Bosphorus database (105/4561), k_{max} + k_{min} detectors.* 120
4.5 *The average number of detected salient vertices on the subsets of expression, pose and occlusion and the whole Bosphorus database.* 123
4.6 *The complementarity and added value of using both k_{max} and k_{min} detectors on the whole Bosphorus database (105/4561).* 124
4.7 *Rank-1 recognition rates with respect to Expression, Pose and Occlusions Variations on the Bosphorus database . CGM is compared with the FGM over the four different local shape descriptors. The meaning and composition of each subset is detailed in Tab. 4.1.* 125
4.8 *Performance comparison on the subset of expressions in the Bosphorus database. The subset frontal encompasses all probe 3D face scans, the subset of occlusion and unlabeled scans included, and only excludes the subset of pose.* 132
4.9 *Performance comparison on the subset of the Bosphorus database with various pose variations.* 132
4.10 *Performance comparison on the subset of the Bosphorus database with diverse occlusions.* 132
4.11 *Performance comparison on the whole Bosphrous database.* 133
4.12 *Robustness to Expression Variations on FRGC v2.0.* 134
4.13 *Performance comparison under expression variations on the FRGCv2 database.* 137
4.14 *Performance comparison on the whole FRGC v2.0 database.* 137
4.15 *Average computation time consumed per salient vertex.* 142

5.1 *Symbol list* 158
5.2 *The numbers of vertices and triangles of the sequence of meshes $\{T_n\}$ with different resolutions.* 167

D.1 Average confusion matrix obtained by HoG 214
D.2 Average confusion matrix obtained by normal cycle based HoS 214
D.3 Average confusion matrix obtained by cubic fitting based HoS 215
D.4 Average confusion matrix obtained by HoG+HoS descriptor using a linear kernel with the SVM 215

D.5 Comparison of the proposed method with the state of the art [11], [9], [5], [6], [8]. 216

E.1 Average, Minimal and Maximal recognition rates obtained by SRC, SVM and SimpleMKL, kernel I: linear; kernel II: RBF; kernel III: chi-square. . . 222
E.2 Average confusion matrix obtained by SimpleMKL-kernel III 222
E.3 Comparison of average recognition rates, (**I: Setup [Gong *et al.* 2009]; II: Setup [Berretti *et al.* 2010b]).** 223

F.1 Confusion matrix: one random test based on MCI. 235
F.2 Average accuracy: MCI and CFI with SRC classifier. 237
F.3 Comparison of 3D FER frameworks and performance on the BU-3DFER dataset (L-Automatic: the framework needs leaning or training a model; F-Automatic: the framework is fully automatic and learning free.) 237
F.4 Average confusion matrix: MCI+CFI 238

G.1 Gallery and probe sets for cases I, II, III, and IV. "A Smile, B Neutral" means that the set contains all images with Twin A smiling and Twin B neutral [Vijayan *et al.* 2011b]. 249
G.2 Comparison of feature level and score level fusion based comprehensive representation (tested in Case I)) 251
G.3 Comparison of rank-one scores of our method (Alg. 5) and the state-of-the-art methods. 251
G.4 Comparison of TAR at 0.1% FAR of our method (Alg. 5) and the state-of-the-art methods. 253
G.5 Comparison of EER of our method (Alg. 5) and the state-of-the-art methods. 254

List of Figures

1.1 *Examples of biometric and non-biometric traits.* . . . 3
1.2 *The picture of a biometric system used for identification or verification.* . . 4
1.3 *Examples of current devices used for capturing fingerprint, iris, and face.* . 5
1.4 *Comparison of MRTD compatibility for various biometric features (Reproduced from [Hietmeyer 2000]).* . . . 6
1.5 *Unlike current machine based systems, human observers are able to handle significant degradations in face images. For instance, subjects are able to recognize more than half of all familiar faces shown to them at the resolution depicted here. The individuals shown from left to right, are: Prince Charles, Woody Allen, Bill Clinton, Saddam Hussein, Richard Nixon and Princess Diana [Sinha et al. 2006].* . . . 7
1.6 *This is a representation of what crime victims often perceive: the threatening weapon in vivid detail, but only a vague sense of the details of the perpetrator. This phenomenon is called "weapon focus" [Sinha et al. 2006]* . . . 7
1.7 *2D face recognition processing flow. Figure is reproduced from [Li & Jain 2005].* . . . 8
1.8 *Milestone approaches in 2D face recognition.* . . . 8
1.9 *Expressive facial expression variations of George W. Bush.* . . . 10
1.10 *Comprehensive examples of head pose variations of Obama.* . . . 10
1.11 *Examples of the facial occlusion variations of Gaddafi.* . . . 11
1.12 *The same face looks totally different under different illuminations: 14 facial illumination variations (columns) under 9 poses (raws) (from the Yale Face Database).* . . . 11
1.13 *Example of a subject in the YouTube Makeup (YMU) database, depicts the subject without [(a)] and with heavy makeup [(b)]; and example of a subject in the Virtual Makeup (VMU) database, (c) depicts the subject without makeup, whereas, (d), (e) and (f) constitute the makeup shots. The makeup shots include the synthetic addition of (d) eye makeup (e) lipstick, and (f) full makeup, respectively, using the popular Taaz software (this figure is assembled from [Dantcheva et al. 2012])* . . . 11
1.14 *From left to right, similarity of faces between twins, Obama and his imitator, a father and his son.* . . . 12
1.15 *From left to right: the photos of Dzhokhar Tsarnaev's the drivers license, student ID card, and images published by the FBI and Massachusetts law enforcement during the manhunt for him and his brother.* . . . 12
1.16 *An example of structured light. A structured light (left), and its projection on to a face. Figure is collected from [Tsalakanidou et al. 2005]* . . . 14
1.17 *Left: Face model captured using a seven camera studio setup; Center: capture systems; Right: Face model captured using consumer binocular stereo camera. Figure is collected from [Beeler et al. 2010]* . . . 15

1.18 *Illustration of the photometric stereo method. From left to right: raw image set captured under four lights; estimated field of surface normals (channels x, y, z respectively); recovered depth map by performing integration on surface normals. Images are collected from [Hansen et al. 2010]* 15

1.19 *Four kinds of 3D face recording formats, and from left to right: texture 3D face face, range image, triangle mesh and the zooming nose region, point clouds and the zooming nose region.* 16

1.20 *3D face recognition processing flow.* 17

1.21 *The images of two twins taken in a session: the top row shows the first twin and the bottom row, the second; the texture images are brightened to increase visibility in this figure.* 19

1.22 *Illustration the examples of exaggerated facial expressions with open mouth. Images taken from the Bosphrous 3D face database.* 19

1.23 *Examples of "nose expressions" sampled from the Bosphorous database.* . . 20

1.24 *An failed example of pose normalization: ICP registration algorithm is performed on a frontal and a partial face with $+90^\circ$ yaw rotation. (a,c): a frontal face scan; (b,d): left profile scan (e): manually pose normalized frontal view of (d); (f): initial position between (c) and (e); (g): (c) and (e) are not correctly registered by the ICP algorithm.* 20

1.25 *Various facial occlusions sampled from the UMB database: e.g. hands, scarves, hats, and other types of occlusion which can occur in real-world scenarios.* 21

1.26 *Illustration of a low-quality data: (a): a pose normalized partial face scan; (b): triangle mesh of (a); (c): the zooming of the region in the red box of (b). We can see that the sampling density of this mesh is very low, and there are many "skinny" triangles on the mesh.* 21

2.1 *Pipleline of the thin-plate-spline based deformation modeling approach [Lu & Jain 2008].* 30

2.2 *Hierarchical surface sampling. (a) first layer (fiducial set); (b) second layer; (c) third layer; (d) final landmark set.* 31

2.3 *Annotated face model and its fitting process: (a) Anthropometric landmarks for face annotation; (b) Annotated facial areas; (c) Checkerboard texture to demonstrate the global UV parameterization; AFM after (d) 0, (b) 8, (c) 32, and (d) 64 iterations. Figures are taken from [Passalis et al. 2005] and [Kakadiaris et al. 2007a].* 33

2.4 *Geometry of a deformable superquadric.* 34

2.5 *Illustration of 3D landmark detection: (a) 8 points based Facial Landmark Model; (b) landmark model overlaid on a 3D facial surface; (c) extrema values of shape index; (d) spin image templet based filter; (e) landmark model fitting for consistent landmark selection; (f) selected best landmarks. Figures are taken from [Perakis et al. 2009] and [Passalis et al. 2011].* . . . 37

2.6 *Illustration of symmetric deformable model fitting: (a) left and right sides of the control mesh of AFM; (b) left and right sides of the annotated AFM; (c)-(e) raw data (top row) and fitted AFM (bottom row) of a subject with different poses. (c) both sides of the frontal scan are fitted; (d) left side of the left scan is fitted (and the right is mirrored); (e) right side of the right scan is fitted (and the left is mirrored). Figures are taken from [Perakis et al. 2009] and [Passalis et al. 2011].* 38
2.7 *Pipeline of training EDM. [Al-Osaimi et al. 2009]* 39
2.8 *Some eigenvectors which span the shape residue PCA subspace (expression deformation model). the images from left to right are the 1st, 2nd, 3rd, 9th, 10th, 11th and 25th.eggenvectors, respectively. Notice that the eigenvectors with the higher eigenvalues are more relevant to the largely deformable regions (e.g. the mouth) and the other regions (e.g. eyebrows, eyes) gradually appear in the lower eigenvectors.[Al-Osaimi et al. 2009]* . . 39
2.9 *Columns 1-3: Expression normalization of two scans of the same subject. The fitting algorithm gives a good estimate (b) of the true face surface given the noisy measurement (a). The pose and expression normalized faces (c) are used for face recognition. Columns 4-6: The reconstruction (b) is robust against scans (a) with artifacts, noise, and holes [Amberg et al. 2008].* 41
2.10 *Main procedures of the Bilinear Model based 3D face recognition [Mpiperis et al. 2008a].* 42
2.11 *Examples of canonical forms (f)-(j) of faces with strong facial expressions (a)-(e) [Bronstein et al. 2005].* 44
2.12 *Facial curves based surface representations: iso-depth curves [Samir et al. 2006], circular curves [Samir et al. 2009] and radial curves [Drira et al. 2013] (from left to right).* 46
2.13 *Illustration of the main procedures for the iso-geodesic strips based approaches: (i) Each face is experimentally partitioned into 9 iso-geodesic stripes, and each strip is further partitioned into 3 parts: lower (L), upper-left (UL) and upper-right (UR) with respect to the coordinate of nose tip. In this figure, the UL parts are expressed by p_1 to p_9 for the left face and g_1 to g_9 for the right face. (ii) The inter-strip distances (i.e. $w(p_i,p_j), i \neq j$, $w(g_i,g_j), i \neq j$) and the intra-strip distances (i.e. $w(p_i,p_j), i = j$, $w(g_i,g_j), i = j$) are computed by using 3D Weighted Walkthroughs. (iii) Faces are compactly represented as connected graphs. (iv) Matching two faces by matching two graphs. The figure is taken from [Berretti et al. 2010a].* 48
2.14 *Extraction of a gallery region and three probe regions. (a) A gallery region, (b) probe C (probe region in general center face area), (c) probe N (probe region in a nose region), (d) probe I (probe region in an interior nose region) [Chang et al. 2006].* 51
2.15 *left: Image of probe sphere centroids labeled by region number. Multiple region numbers at a centroid indicate that more than one radius was used for cropping, yielding multiple region probes with the same centroid. (a) 04514d324.* right: *Examples of nose regions with different radii [Faltemier et al. 2008a].* 52

2.16 *Overview of the regional registration based 3D face recognition approach [Alyuz et al. 2010].* . . . 53
2.17 *The main stages of the Simulated Annealing based 3D face recognition framework and the main stages of the Simulated Annealing based face matching (in the red box) [Queirolo et al. 2010].* . . . 55
2.18 *Interpenetrating points p in A with respect to B [Silva et al. 2005].* . . . 55
2.19 *30 local regions used for face comparison[Spreeuwers 2011].* . . . 56
2.20 *Rank-one scores of individual region classifiers on the FRGCv2 database.* . . . 57
2.21 *Decomposition of the face into subregions.* . . . 58
2.22 *The framework of Wang* et al.*'s method [Wang et al. 2010].* . . . 59
2.23 *MS-eLBP-DFs of a facial range image with different radii from 1 to 8 (from left to right), the layer 1 is the same as original LBP, from layer 2 to layer 4, the absolute difference between central pixel and its neighbors are encoded [Huang et al. 2012].* . . . 60
2.24 *From left to right: segmentation of two faces by sign of Gaussian and mean curvature: convex (white), concave (black), and two types of saddle: positive mean curvature (light grey), negative mean curvature (dark grey) [Gordon 1992]. maximum principle directions; minimum principle directions; rigid lines; valley lines; EGI of rigid lines (top), and EGI of valley lines (bottom) [Tanaka et al. 1998].* . . . 61
2.25 *Facial mask (left), An original face mesh and its Remeshings [Li et al. 2009].* 62
2.26 *Illustration of three local shape descriptors: spin image (left), point signature (middle), and local shape map (right).* . . . 62
2.27 *Illustration of three local shape descriptors: 3D shape context (left), spherical face representation (middle), and tensor-based face representation (right).* . . . 63

3.1 *Illustration of patch weights for 2D and 3D face recognitions: (a-b) a 2D face image and its corresponding patch weights [Ahonen et al. 2004a]; (c-d) a 3D face depth image and its corresponding patch weights learned by our method. All images are split to 6×6 local patches. Darker patches indicate lower weights, while brighter ones indicate higher weights.* . . . 71
3.2 *Overview of the proposed approach.* . . . 73
3.3 *Illustration of two approaches to normal estimation: (a) a plane is fitted to a vertex p and its neighbors; (b) the normal vectors of triangles in one-ring of p are averaged.* . . . 74
3.4 *Illustration of facial normal estimation: (a) the original range image, (b-d) its normal images of component x, y and z (the sample comes from the 3D-TEC dataset).* . . . 75
3.5 *Examples of the neighborhood of LNP: (a) $Q_{1,8}$ and (b) $Q_{2,16}$.* . . . 76
3.6 *Illustration of facial normal encoding: (a) to (c), normal images of component x, y and z; (d) to (e), their corresponding LNP maps using the neighborhood $Q_{1,8}$.* . . . 76
3.7 *Illustration of facial normal representation: histogram of LNP.* . . . 77
3.8 *Illustrate of the raw samples of the five databases: (a) FRGC v1.0, (b) FRGC v2.0, (c) Bosphorus, (d) BU-3DFE, (e) 3D-TEC.* . . . 80

3.9 *Illustration of several examples of preprocessed face models: first row: models of one subject with different facial expressions (BU-3DFE); second and third rows: models of one subject with different facial expressions and action units (Bosphorus); fourth row: models of one subject with different facial expressions (FRGC v2.0); last row: two pairs of identical twin models with neutral and smiling expressions (3D-TEC).* 82
3.10 *Illustration of the patch weights learned from the Bosphorus database. Columns (a-c), normal images x, y and z and their patch weights (10×8 patches); columns (d-f), normal images x, y and z and their patch weights (6×6 patches); columns (g-i), normal images x, y and z and their patch weights (3×3 patches). Darker patches indicate lower weights, while brighter ones indicate higher weights.* 88

4.1 *2D texture images (1^{st} row) and 3D face scans (2^{nd} row) from the Bosphorus dataset illustrating the various challenges in real-life applications under uncontrolled environment: (a) frontal and neutral; (b) facial expression variation; (c) pose variation and data missing due to self-occlusion; and (d) external occlusion.* 98
4.2 *Overview of the proposed method.* 100
4.3 *Illustration of normal estimation. (a) one-ring neighborhood, (b) $\mathbf{n}_{v_1}$ is computed by averaging the normal vectors of triangles in one-ring of v_1.* . . 107
4.4 *Illustration of multiple order surface differential quantities estimation. (a) original facial surface; (b) normal vector; (c) maximum principle curvature; (d) minimum principle curvature; (e) shape index; (f) gradient of shape index;* 108
4.5 *Illustration of the salient vertex detection: salient vertices detected by κ_{max} detector (a), and κ_{min} detector (b), over a facial surface.* 109
4.6 *Canonical direction(s) assignment: (a) a detected salient vertex and its geodesic disk patch; (b) coordinate transformation and normal vector estimation; (c) normal vector projection and initial direction selection; (d) angle statistics and canonical direction(s) assignment.* 110
4.7 *The configuration of a local shape descriptor : Canonical orientation (red vector stemming from the centre vertex with an arrow), the centre salient vertex (+) and its 8 neighborhood vertices (+) with a distance of r_1; 9 circles with the same radius of r_2 are assigned to the 9 vertices, respectively.* 111
4.8 *2D Samples of 6 basic facial emotions with neutral: (a) neutral; (b) Anger; (c) Disgust; (d) Fear; (e) Happiness; (f) Sadness; (g) Surprise.* 117
4.9 *2D Samples of facial action units: (a-f) displaying Lower Action Units, (g-j) displaying Upper Action Units, and (k-l) representing Combined Action Units.* 117
4.10 *Samples of head pose variations: (a) frontal; (b) Yaw $+10°$; (c) Yaw $+20°$; (d) Yaw $+30°$; (e) Yaw $+45°$; (f) Yaw $-45°$; (g) Yaw $+90°$; (h) Yaw $-90°$; (i) PR: Strong upwards; (j) PR: Strong downwards; (k) CR: $+45°$ yaw and approximately $+20$ pitch; (l) CR: $+45°$ yaw and approximately -20 pitch.* 118
4.11 *Samples of occlusions: from (a) to (d), Occlusion by glasses, hair, by hand on right eye and by hand on mouth.* 118

4.12 *Repeatability of salient vertices on the Bosphorus database.* 122
4.13 *The average number of detected salient points on each subset of the Bosphorus database.* 123
4.14 *The rank-1 recognition rates by the coarse grained matcher (CGM) over the various subsets of the Bosphorus database.* 127
4.15 *The rank-1 recognition rates by the fine grained matcher (FGM) over the various subsets of the Bosphorus database.* 128
4.16 *Comparison of the coarse grained matcher (CGM) and the fine grained matcher (FGM) on the whole Bosphorus database when MV-HOG is used as local shape descriptor.* 129
4.17 *On the usefulness of salient point detector fusion and feature fusion. Early fusion of local shape descriptors: (a) coarse grained matcher (CGM); (b) fine grained matcher (FGM). Score-level fusion of both the κ_{max} and κ_{min} detectors: (c) CGM; (d) FGM.* 130
4.18 *Illustration of performance on expression and pose subsets. (a) performance of coarse-matcher on expression subsets; (b) performance of fine-matcher on expression subsets; (c) performance of coarse-matcher on pose subsets; (d) performance of fine-matcher matcher on pose subsets.* 131
4.19 *Illustration the Cumulative Match Curves on the Bosphorus database.* . . . 133
4.20 *The Cumulative Match Curves on FRGC v2.0.* 135
4.21 *Illustration of the Cumulative Match Curves on the FRGCv2 database.* . . . 136
4.22 *Gaining insights into the proposed local feature-based 3D FR framework. The four cases are displayed using the κ_{max} salient vertex detector and MV-HOMQ/CGM framework: (a) Neutral vs. Neutral; (b) Neutral vs. Expression; (c) Neutral vs. YR; (d) Neutral vs. Occlusion.* 138
4.23 *Examples of non correctly recognized samples with hair occlusion.* 138
4.24 *Examples of non recognized faces with large pose variations (YR90°) in Bosphorus.(a): gallery neutral face scan; (b-c):left profile scan (b) and its frontal view (c); (d-e): right profile scan (d) and its frontal view; (e): manual registration of (d) over (a) showing that more than 50% face data is missing in the right profile scan (d).* 139
4.25 *Illustration of the mesh quality of the sample with large pose variation: (a) the scan of BS001YRR90; (b) corresponding triangulation mesh of (a); (c) the zooming of (b) over the region around the eye and nose.* 139
4.26 *Examples of non correctly recognized samples with disgust expression.* . . . 140
4.27 *Examples of non correctly recognized samples with lower action unit.* . . . 140
4.28 *Illustration of three incorrect matching results: (a) Disgust expression; (b) Extreme pose; (c) External occlusion by hair.* 141

5.1 *Uniformization for closed surfaces.* 151
5.2 *Uniformization for surfaces with boundaries.* 151
5.3 *Offset surface and tube formula.* 154
5.4 *Additivity of the normal cycle.* 155
5.5 *Configuration.* 157
5.6 *Small triangles inscribed to attitudinal circles of a cylinder do not guarantee the normal convergence.* 160

5.7 *Homotopy between the normal cycles $N(T)$ and $N(M)$.* . . . 162
5.8 *Pipeline for meshing a Bézier patch of Utah teapot.* . . . 165
5.9 *Remeshing of the Car spline surface model.* . . . 166
5.10 *The maximal and average circumradii $\{\varepsilon_n\}$ (a-b), and the minimal and average of minimal angles of $\{T_n\}$ (c-d).* . . . 168
5.11 *Curvature errors e_n^G and e_n^H of $\{T_n\}$ converge to zeros as the number of sample points goes to infinity (a-b), and as the average of circumradii $\{\varepsilon_n\}$ goes to zero (c-d).* . . . 168
5.12 *Illustration of the curvature values on the Utah teapot spline surface patch M, (a, d), and on its approximate mesh T_{11} (b, e). Their relative curvature error distribution histograms are shown in (c) and (f).* . . . 169

A.1 *Regular parameterized surface patch S.* . . . 176
A.2 *Illustration of the second fundamental form.* . . . 178
A.3 *Normal curvature.* . . . 178
A.4 *Normal curvature.* . . . 179
A.5 *Gauss map.* . . . 179
A.6 *Smooth manifold.* . . . 182
A.7 *Tangent map.* . . . 184

B.1 *1, 2, 3-simplices in $\mathbb{R}^3$.* . . . 189
B.2 *A non-manifold simplicial complex.* . . . 190
B.3 *Gaussian curvature at point v.* . . . 191
B.4 *Dihedral angle on edge $e = [v_0, v_1]$.* . . . 191
B.5 *Circle packing metric.* . . . 192

C.1 *A Delaunay triangulation in $\mathbb{E}^2$ with circumcircles shown.* . . . 196
C.2 *Any encroached segment is split into subsegments by inserting a new vertex at its midpoint.* . . . 196
C.3 *Insert vertices at circumcenters of triangles with small angles (e.g., $<$ 20.7°).* . . . 197
C.4 *Hexagons inscribed in a circle, $n = 6, \theta = \frac{\pi}{6}$.* . . . 198
C.5 *Illustration of a Schwarz Lantern.* . . . 199
C.6 *Flattening a geodesic triangle derives an estimation of the Gauss curvature.* 200
C.7 *Notations.* . . . 201

D.1 Illustration of normal cycle theory based curvature estimation method (equation (1) [19]). . . . 209
D.2 First row, from left to right: κ_{max}, κ_{min} and shape index estimated by normal cycle theory based method. Second row, from left to right: κ_{max}, κ_{min} and shape index estimated by cubic fitting based method, (model M0044-DI03). 210
D.3 Illustrated the estimated normal vectors (model M0044-DI03). . . . 211
D.4 From left to right, 60 selected manual landmarks, local neighborhood points of the left mouth corner (M0044-DI04 and M0044-DI03). . . . 212
D.5 Canonical orientation (arrow), landmark point (o) and its 8 neighborhood vertices (+) assigned with 9 circles. . . . 212

D.6 Examples of six universal expressions, from left to right: anger, disgust, fear, happiness, sadness, surprise. First row, high intensity, Second row, highest intensity. 213

E.1 MS-LNPs: images of normal components X, Y and Z and corresponding features extracted from three scales: $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$ (from left to right). 219

F.1 Overview of the proposed 3D FER framework. 227
F.2 Conformal images for 3D human facial expressions. The subject is M0021 in BU3D database. 1st row: 6 expression images in camera view; 2nd row: conformal maps with human facial texture; 3rd row: mean curvature images; 4th row: conformal factor images. 230
F.3 Example of right expression classification: (a) a testing MCI with disgust expression; (b) the reconstructed MCI with minimal reconstruction error; (c) the minimal reconstruction error image (i.e. the difference of (a) and (b)); (d) the sparse representation coefficients over MCI based expression dictionary (the range of disgust dictionary is from 109 to 216); (e) the expression-class-dependent reconstruction errors. 234
F.4 Example of wrong expression classification: (a) a testing MCI with disgust expression; (b)-(e) have the same meanings as the ones in Fig. F.3. 234
F.5 72 randomly selected testing MCIs: 6 subjects and 6 classes expressions. The samples of two adjacent columns belongs to the same subject and the samples of each row belongs to the same expression class. The wrong classified samples are indicated by red boxes. 235
F.6 72 reconstructed MCIs with the minimal errors corresponding to Fig. F.5. The wrong classified samples are indicated by red boxes. 236
F.7 The minimal expression-class-dependent reconstruction errors corresponding to Fig. F.5 and F.6. 236

G.1 Framework of our proposed method. 242
G.2 From left to right: the original range image, and its normal images of component x, y and z. 244
G.3 (a) to (c): normal images of component x, y and z; (d) to (e), the corresponding N-LBPs (LNPs) representation $Q_{1,8}$. 245
G.4 (a). Normal component z. (b) Its Gabor representation. 246
G.5 (a). Normal component z. (b) Its LGBP representation. 247
G.6 ROC curves in four cases of our method based on N-LBPs. The legend shows TAR at 0.1% FAR. 252
G.7 ROC curves in four cases of our method based on N-GFs. The legend shows TAR at 0.1% FAR. 253
G.8 ROC curves in four cases of our method based on N-LGBPs. The legend shows TAR at 0.1% FAR. 254

Abstract

Due to the natural, non-intrusive, easily collectible, widespread applicability, machine-based face recognition has received significant attention from the biometrics community over the past three decades. Compared with traditional appearance-based (2D) face recognition, shape-based (3D) face recognition is more stable to illumination variations, small head pose changes, and varying facial cosmetics. However, 3D face scans captured in unconstrained conditions may lead to various difficulties, such as non-rigid deformations caused by varying expressions, data missing due to self occlusions and external occlusions, as well as low-quality data as a result of some imperfections in the scanning technology.

In order to deal with those difficulties and to be useful in real-world applications, in this thesis, we propose two 3D face recognition approaches: one is focusing on handling various expression changes, while the other one can recognize people in the presence of large facial expressions, occlusions and large pose various. In addition, we provide a provable and practical surface meshing algorithm for data-quality improvement.

To deal with expression issue, we assume that different local facial region (e.g. nose, eyes) has different intra-expression/inter-expression shape variability, and thus has different importance. Based on this assumption, we **design a learning strategy** to find out the quantification importance of local facial regions in terms of their discriminating power. For facial description, we **propose a novel shape descriptor** by encoding the micro-structure of multi-channel facial normal information in multiple scales, namely, **Multi-Scale** and **Multi-Component Local Normal Patterns** (MSMC-LNP). It can comprehensively describe the local shape changes of 3D facial surfaces by a set of LNP histograms including both global and local cues. For face matching, Weighted Sparse Representation-based Classifier (W-SRC) is formulated based on the learned quantification importance and the LNP histograms. The proposed approach is evaluated on four databases: the **FRGC v2.0, Bosphorus, BU-3DFE and 3D-TEC**, including face scans in the presence of diverse expressions and action units, or several prototypical expressions with different intensities, or facial expression variations combine with strong facial similarities (i.e. identical twins). Extensive experimental results show that the proposed 3D face recognition approach with the use of discriminative facial descriptors can be able to deal with expression variations and perform quite accurately over all databases, and thereby has a good generalization ability.

To deal with expression and data missing issues in an uniform framework, we **propose** a mesh-based registration free 3D face recognition approach based on a novel local facial shape descriptor and a multi-task sparse representation-based face matching process. The proposed approach consists of three main steps: 1) A set of facial salient points are detected for each facial surface by using the local extreme values of principle curvatures in 3D Gaussian scale space; 2) To describe each salient point associated with a local support region, **we propose** to compute the weighted statistical distributions of multiple order surface differential quantities, including Histogram of mesh Gradient (HOG), Histogram of Shape index (HOS) and Histogram of Gradient of Shape index (HOGS); 3) Differen-

t from the coarse grained matching scheme which simply counts the number of matched salient points in a pair of face scans, **we propose** a fine grained matching scheme by solving a multi-task sparse representation problem. Comprehensive experiments were carried out on Bosphorus, the largest public database which best accounts for the various challenges of 3D face recognition in uncontrolled conditions, i.e. expression variations, pose changes and external occlusions. The experimental results demonstrate the effectiveness of the proposed method which achieves the best state of the art rank-1 recognition rates over the whole dataset (**96.56%**) and its three subsets of expression variations (**98.82%**), pose changes (**91.14%**) and occlusions (**99.21%**). Additional experiments on the FRGC v2.0 database show that the proposed method consistently displays competitive performance in comparison with state of the art algorithms which mostly require accurate face alignment and/or learning of discriminating features, and thereby demonstrates its generalization ability.

The direct motivation to deal with the low-quality data issue in this thesis is to achieve reliable facial surface descriptions. It is well known that the estimates of surface differential quantities such as surface normal and curvatures are very sensitive to noise and mesh quality. To handle this issue, we **present** a conformal mapping-based surface meshing algorithm to generate high-quality triangle meshes with theoretic guarantee of surface normal and curvature convergence. Specifically, given a smooth surface embedded in $\mathbb{E}^3$, our surface meshing algorithm can generate a sequence of well-shaped and well-sized triangulations by performing classical 2D Delaunay refinement algorithms on conformal uniformization domain. To compare the local geometry of the smooth surface and its discrete approximations, we **choose to use** the normal cycle theory, which gives a general framework involving both smooth and discrete curvatures. This implies that we deal with curvature measures, which allow to compare the local curvatures of smooth surface and triangulation. We **give** explicit error estimates and **prove** the convergence of surface normal and curvature measures. We evaluated our theoretical results by meshing smooth Spline surfaces. The experimental results verify the theoretical results and demonstrated the efficiency of the designed surface meshing algorithm. More experiments for 3D face recognition with data-quality guarantee will be tested in our future work.

Keywords: 3D face recognition, expressions, occlusions, and pose variations, facial surface descriptors, surface meshing, normal and curvature convergence.

Résumé

En raison des naturelle, non-intrusive, facilement percevable caractéristiques, et une large diffusive applicabilité pour la criminalistique et de la sécurité, reconnaissance faciale basée sur la machine a reçu beaucoup d'attention de la communauté biométrie au cours des trois dernières décennies. Par rapport à la traditionnelle reconnaissance faciale basée sur le visage 2D, la recconnaissance faciale basé sur la forme 3D est plus stable aux variations d'éclairage; petite changements de tête pose, et variant cosmétiques pour le visage. Cependant, le visage 3D numérise capturé dans des conditions non-contraintes peut conduire à des difficultés diverses, comme des déformations non rigides provoquées par la variant expressions, les données manquantes en raison de l'auto-occlusions et des occlusions externes, ainsi que des données de faible qualité en raison de certaines imperfections de la technologie de numérisation. Pour régler ces difficultés et d'améliorer les applications du monde réel, dans cette thèse, nous proposons deux approches de 3D reconnaissance faciale: l'un se concentre sur le handling de divers changements d'expression, l'autre peut reconnaître les gens à la situation de présence d'un grand les expressions facial, des occlusions et des grands pose divers. En outre, nous fournissons une surface prouvable et pratique algorithme de surface maillage pour l'amélioration de la qualité de données.

Pour faire face aux problème d'expression, nous supposons que la variabilité des formes de intra-expression/inter-expression de la faciale local région différent (e. g., nez, yeux) est différent, et a donc une importance niveau différente. Sur la base de cette hypothèse, nous concevons une stratégie d'apprentissage pour découvrir l'importance de la quantification des régions faciales locaux en fonction de leur énergie discriminant. Pour une description du visage, nous proposons une nouvelle descripteur pour coder la microstructure du multi- canal d'information normale du visage dans multiples échelles, à savoir, Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP). On peut globalement décrire les changements de forme locale de 3D surfaces faciales par un ensemble d'histogrammes LNP y compris les indices globaux et locaux. Pour le visage correspondant, Weighted Sparse Representation-based Classifier (W-SRC) est formulée sur la base de l'importance de la quantification appris et les histogrammes LNP. L'approche proposée est évaluée sur quatre bases de données: le FRGC v2. 0, Bosphore, BU- 3DFE et 3D -TEC, y compris les scans du visage en présence de diverses expressions et des unités d'action, ou de plusieurs expressions prototypiques avec des intensités différentes, ou des variations d'expression du visage combinée avec de fortes similitudes faciales (c.à.d. jumeaux identiques). Résultats expérimentaux étendus montrent que l'approche de reconnaissance de 3D visage proposé avec l'utilisation de descripteurs discriminants du visage peut régler les variations d'expression et d'effectuer avec assez de précision sur toutes les bases de données, et a ainsi une bonne capacité de généralisation.

Pour faire face à l'expression et problème des données manquantes dans un cadre uniforme, nous proposons une approche sur le sans-enregistrement maillage-basé reconnaissance du 3D visage basé sur un nouveau local descripteur de la forme du visage et un correspondance processus d'clairsemée représentation du visage basée multi- tâche. L'approche

proposée consiste en trois étapes principales: 1) Un ensemble de points saillants du visage sont détectés pour chaque surface du visage en utilisant les valeurs extrêmes locales de courbures de principe dans le 3D Gaussienne échelle espace, 2) Pour décrire chaque point saillant associé à un local région supporté, nous proposons de calculer les distributions pondérées statistiques de plusieurs échelles du surface de quantités différentielles, y compris Histogram of mesh Gradient (HOG), Histogram of Shape index (HOS) and Histogram of Gradient of Shape index (HOGS); 3) Ce qui différecie de le grossière grain appariement qui comptes simplement le nombre de points saillants correspondants dans une paire de scans du visage, nous proposons un schéma correspondant avec grain fin en résolvant un multi-tâche problème de représentation clairsemé. Expériences exhaustives ont été effectuées sur le Bosphorus, ce qui est la plus grande base de données publique qui est le meilleur des différents enjeux de la reconnaissance faciale en 3D dans des conditions non-contrôlées, soit des variations d'expression, changements de pose et les occlusions externes. Les résultats expérimentaux démontrent l'efficacité de la méthode proposée, qui réalise le meilleur état de l'art de rang 1 des taux de reconnaissance sur l'ensemble des données (96.56%) et de ses trois sous-ensembles de variations d'expression (98.82%), changement de pose (91.14%) et occlusions (99.21%). Des expériences complémentaires sur base de données FRGC v2.0 presentes que la méthode proposée affiche constamment la performance compétitive en comparaison avec l'état de l'art des algorithmes qui demandent surtout de l'alignement de visage précis et/ou l'apprentissage des caractéristiques distinctives, et démontre ainsi sa capacité de généralisation.

La motivation directe de faire face à la problème des données de faible qualité dans cette thèse est de gagner des descriptions de surface faciale fiables. Il est bien connu que les estimations des quantités différentielles de surface tels que la normale de surface et des courbures sont très sensibles à la qualité de maillage. Pour traiter cette problème, nous présentons une surface cartographie-basée conforme algorithme de maillage pour générer des maillages triangulaires de haute qualité avec garantie théorique de la normale de la surface et de la convergence de courbure. Plus précisément, étant donné une surface lisse intégré dans l'espace euclidien à trois dimensions, notre maillage surface algorithme peut générer une séquence de triangulations bien formées et bien dimensionné en exécutant des algorithmes de raffinement de Delaunay 2D classiques sur le domaine de l'uniformisation conforme. Pour comparer la géométrie locale de la surface lisse et ses approximations discrètes, nous choisissons d'utiliser la théorie du cycle normal, ce qui donne un cadre général impliquant deux courbures douces et discrètes. Cela implique que nous traitons avec des mesures de courbure, ce qui nous permet de comparer les courbures locales de surface lisse et triangulation. Nous donnons des estimations d'erreur explicites et de prouver la convergence de la normale de la surface et les mesures de courbure. Nous avons évalué nos résultats théoriques par maillage de splines surfaces lisses. Les résultats expérimentaux vérifient les résultats théoriques et ont démontré l'efficacité de le maillage de surface destinée algorithme. D'autres expériences de reconnaissance 3D faciale avec une garantie de qualité des données seront testées dans nos travaux futurs.

Mots-clés: reconnaissance de visage en 3D, expressions, des occlusions et poser variations, descripteurs de surface faciale, maillage surfacique, la convergence de la normale et la courbure.

CHAPTER 1

Introduction

Contents

1.1 Context and Motivation 1
1.1.1 What is Biometrics? 1
1.1.2 Why Face Recognition? 4
1.1.3 Face Recognition: A Brief Review in 2D 6
1.1.4 Face Recognition: Challenges in 2D 9
1.1.5 Face Recognition: Opportunities in 3D 13
1.2 3D Face Recognition 13
1.2.1 3D Face Acquisition Techniques 13
1.2.2 Basic Concepts and Terminology 15
1.2.3 3D Face Databases 17
1.2.4 Main Challenges and Objectives 18
1.3 Thesis Contributions 21
1.3.1 Methodologies 21
1.3.2 Main Contributions 22
1.4 Thesis Organization 27

1.1 Context and Motivation

1.1.1 What is Biometrics?

Generally speaking, "Biometrics" means "life measurement", which is the science and technology of measuring and analyzing biological data. But the term is usually used as the same meaning of "Biometric Authentication" or "Biometric Recognition". Thus, when we are talking about "Biometrics", it usually means the science and technology of the identification of humans by measuring (e.g. aligning, describing and comparing *etc.*) their life related, physiological, chemical, or behavioral attributes, characteristics or traits.

The selection of a particular trait for use in a specific application involves a weighting of several factors. Jain *et al.* [Jain *et al.* 1999] identified seven such factors to be used when assessing the suitability of any trait for use in biometrics.

1. *Universality:* Every person using the system should possess the trait.

2. *Uniqueness:* The trait should have sufficient unique properties to distinguish one individual from any other.

3. *Acceptability:* The capturing technology should be invasiveness-free and possible in a way acceptable to a large percentage of the relevant population.

4. *Measurability:* The trait should be easy for capturing and suitable for measuring.

5. *Permanence:* A trait with "good" permanence will be reasonably invariant over a long period of time.

6. *Circumvention:* The trait should be difficult to be imitated using an artifact or substitute.

7. *Performance:* It relates to the accuracy, speed, and robustness of technology used involving the trait.

Among the various biometric technologies being considered, the traits which satisfy the above requirements are fingerprint, face, iris, hand geometry, ear shape, palm print, voice, retina, vein patterns, DNA, odor, signature, keystroke dynamics, *etc.* In addition, some very common human attributes (e.g. gender, age, ethnicity, height, hair color, *etc.*), which are shared by a large group of population and can be used for coarsely clustering and labeling humans into subgroups, are called "soft biometric traits".

In contrast, non-biometric traits, including token-based documents (e.g. a driver's license or a person's passport) and knowledge-based codes (e.g. a password or personal identification number), have a large probability of being lost, forgotten, stolen, counterfeited, damaged, *etc.* Please refer to Fig. 1.1 for the examples of biometric and non-biometric traits.

A biometric system can be either an "identification (recognition)" system or a "verification (authentication)" system, which are defined as follows.

- **Identification** It performs a one-to-many matching between a new captured biometric sample and the enrolled biometric templates stored in a biometric database in attempt to determine the identity of an unknown individual.

- **Verification** It performs a one-to-one matching between a new captured biometric sample and a specific biometric template stored in a biometric database in attempt to verify the individual is the person he or her claims to be.

The main picture of both identification and verification systems involves in three main steps, including data capture and preprocessing, feature extraction and template generation, and templates matching (see Fig. 1.2).

Considering a biometric system involves a special group of population, for the first time, their biometric information associated with their identify information should be captured and stored in the biometric system, which is called *enrollment*. In the first step, the desired biometric trait of a user is captured by a specific sensor such as a digital camera. The following preprocessing process ususlly includes artifacts and noises removing, pose aligning, scale normalization, *etc.* In the second step, the system performs feature extraction to generate biometric template (a mathematical representation of the biometric trait). If enrollment is being performed, the template is simply stored somewhere (on a card or within a database or both). Finally, the new template is passed to a mater that compare

Figure 1.1: *Examples of biometric and non-biometric traits.*

it with other existing templates, estimating the distance between them using any metric (e.g. Euclidean distance). The matcher will output its decision which can be used for any specified application such as border entry or exit.

The following metrics are used for the standard performance measurement of a biometric system with a closed-set identification scenario.

- **Rank-one Recognition Rate** The rate that the system correctly matches the input template to a matching template in the database with a highest similarity. It measures the percent of inputs which are correctly identified.
- **Cumulative Match Characteristic (CMC)** The similarities between the input template and all the templates in the database are first ranked. Then, the CMC curve shows how often the individual's template appears in the ranks (1, 5, 10, 100, etc.), based on the match rate.

The following metrics are used for the standard performance measurement of a biometric system with a verification scenario.

- **False Accept Rate (FAR)** The rate that the system incorrectly matches the input template to a non-matching template in the database. It measures the percent of invalid inputs which are incorrectly accepted.
- **False Reject Rate (FRR)** The rate that the system fails to detect a match between the input template and a matching template in the database. It measures the percent of valid inputs which are incorrectly rejected.

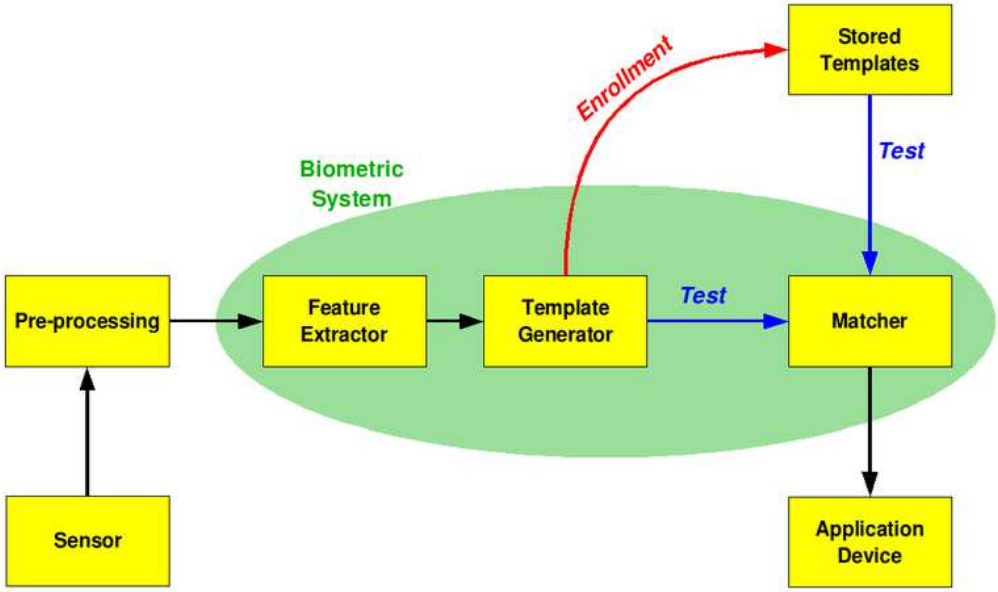


Figure 1.2: *The picture of a biometric system used for identification or verification.*

- **Receiver Operating Characteristic (ROC)** The ROC curve is a visual demonstration of the trade-off between the FAR and the FRR related to a varying threshold.
- **Equal Error Rate (EER)** A point of ROC curve at which FAR equals to FRR. The EER is a quick way to compare the accuracy of systems with different ROC curves. In general, the system with the lower EER is more accurate.

As pointed out in the book "Handbook of Biometrics" [Anil K. Jain 2007], biometrics is a rapidly evolving field with widely applications ranging from securely accessing one's computer to gaining entry into a country. The basic concepts introduced in the above section is referred from the web of Wikipedia in terms of "Biometrics". For more information, please refer to the handbooks [Anil K. Jain 2007] [Anil K. Jain 2011].

1.1.2 Why Face Recognition?

Given the requirement for determining people's identity, the obvious question is what is best suited to supply this information [Pentland & Choudhury 2000]? Among all the available biometric traits, fingerprint, iris, face, voice and signature have gained widely attention. In particular, fingerprint, iris and face recognition techniques grow mature gradually, and become the mainstream biometric recognition technologies so far. Although fingerprint and iris recognition techniques can achieve reliable performance in some particular commercial applications, they have the following obvious disadvantages:

1. *Physically intrusive:* Fingerprint requires user's cooperation in making physical contact with the sensor's surface, while current iris imaging requires user to position their eye relative to the sensor (see Fig. 1.3). Both facts raise problems of how to keep the surface clean and germ-free in a high-throughput application [Bowyer *et al.* 2006]. Moreover, this kind of cooperation scanning techniques also

Figure 1.3: *Examples of current devices used for capturing fingerprint, iris, and face.*

require user to pause for a second to 'declare' themselves. This 'pause and declare' interaction is unlikely to change because of the fine-grain spatial sensing required [Pentland & Choudhury 2000].

2. *Socially intrusive:* There is a 'oracle-like' aspect to the interaction: since people can't recognize other people using this sort of data, these types of identification do not have a place in normal human interactions and social structures [Pentland & Choudhury 2000].

While the 'pause and present' interaction and the oracle-like perception are useful in high-security applications (they make the systems look more accurate), they are exactly the opposite of what is required when building a store that recognizes its best customers, or an information kiosk that remembers you, or a house that knows the people who live there [Pentland & Choudhury 2000]. Face recognition from video and voice recognition have a natural place in these next-generation smart environments,

1. *Natural and non-intrusive:* They are unobtrusive (able to recognize at a distance without requiring a 'pause and present' interaction), are usually passive (do not require generating special electro-magnetic illumination), do not restrict user movement, and are now both low-power and inexpensive [Pentland & Choudhury 2000]. See Fig. 1.3 for some examples of face capturing devices.

2. *Biological perception:* Perhaps most important, however, is that humans identify other people by their face and voice, therefore are likely to be comfortable with systems that use face and voice recognition [Pentland & Choudhury 2000].

Similar superiorities are also pointed by Hietmeyer [Hietmeyer 2000], in which he claims that "Facial recognition is non-intrusive and does not require a person to perform an overt, time-consuming act to allow details to be captured, nor does it require touching something. The facial image can be captured repeatedly, even from a distance".

As a strong evidence from the International Civil Aviation Organization (ICAO), Hietmeyer [Hietmeyer 2000] pointed out that biometric identification can promise fast and secure processing of airline passengers. To select a single biometric trait for international

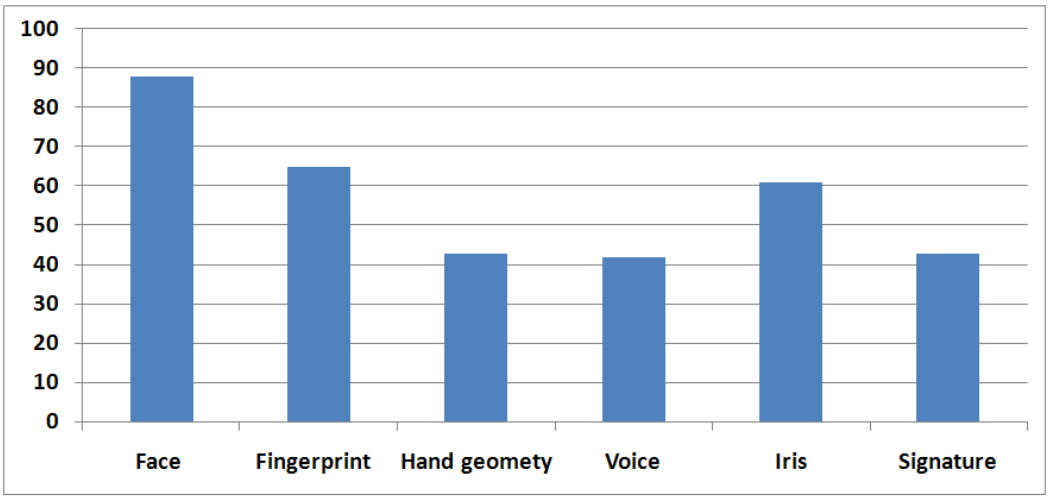


Figure 1.4: *Comparison of MRTD compatibility for various biometric features (Reproduced from [Hietmeyer 2000]).*

use in machine-assisted identity confirmation, they reported the compatibility evaluation of six biometric traits including face, fingerprint, hand geometry, voice, eye as well as signature based on the Machine Readable Travel Documents (MRTD) system. The compatibility score involves the factors of enrollment, renewal, machine-assisted identity verification requirements, redundancy, public perception, storage requirements and performance. As shown in Fig. 1.4, face recognition system scored the highest compatibility and is emerging as the most likely biometric to be select for international use. Fingerprint and iris followed and compatibility of signature, hand geometry and voice scored last.

1.1.3 Face Recognition: A Brief Review in 2D

Face recognition is an inherent ability of humans, and a nature way of people knowing and remembering each other in their daily life. Benefiting of human visual system, humans are extremely good at recognizing faces they are familiar with and can do so despite nature changes in expression, viewpoint and hairstyle as well as image manipulation such as inversion [Pike *et al.* 2000]. One of the interesting findings by Sinha *et al.* [Sinha *et al.* 2006] concludes that humans can recognize faces in extremely low-resolution images (Fig. 1.5). Wilmer *et al.* presented an unusual phenomenon: human face recognition is a highly specific and highly heritable cognitive ability [Wilmer *et al.* 2010]. However, humans also have their own limitations for face recognition. Several evidences have been found that human memory for briefly seen faces is rather poor [Pike *et al.* 2000] [Sinha *et al.* 2006]. For example, eyewitnesses to a crime find it very hard to describe and particularly to identify the perpetrator and are very susceptible to exposure to similar faces and the instructions given to them [Pike *et al.* 2000]. A reasonable explanation is that crime victims are typically terrified of weapons being used by the perpetrator, making them much less aware of the details of the face of the perpetrator. This kind of phenomenon is called "weapon focus" (illustrated in Fig. 1.6) . It is obviously that machine-based face recognition systems can immune to such emotional influences, and have the chances of exceeding human

Figure 1.5: *Unlike current machine based systems, human observers are able to handle significant degradations in face images. For instance, subjects are able to recognize more than half of all familiar faces shown to them at the resolution depicted here. The individuals shown from left to right, are: Prince Charles, Woody Allen, Bill Clinton, Saddam Hussein, Richard Nixon and Princess Diana [Sinha et al. 2006].*

Figure 1.6: *This is a representation of what crime victims often perceive: the threatening weapon in vivid detail, but only a vague sense of the details of the perpetrator. This phenomenon is called "weapon focus" [Sinha et al. 2006] .*

performance under some stressful circumstances.

As a typical biometric system shown in Fig. 1.2, machine-based 2D face recognition system consists of four modules as describe in Fig. 1.7: face detection and alignment for face preprocessing, as well as feature extraction and matching for face recognition. Given an image or section of video captured from a face sensor, *Face detection* segments out the face regions from their background using a *face detector*. In the case of video, the detected faces may also need to be tracked by a *face tracker*. *Face alignment* aims to align all the detected faces in the same position. Meanwhile, scale, pose and illumination normalization are also necessary components. The normalized faces are used for facial *feature extraction*, which usually provides the discriminating representations of faces as feature vectors. In the following *face matching* step, each feature vector of a testing face is match against those of enrolled faces in the database. Finally, the system outputs the identity of the face when a match is found with sufficient confidence.

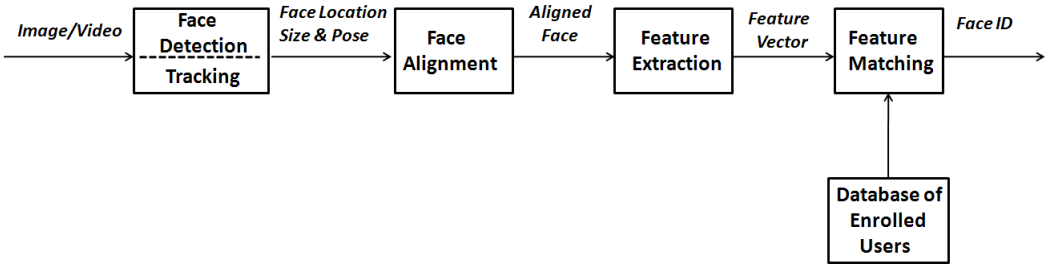


Figure 1.7: *2D face recognition processing flow. Figure is reproduced from [Li & Jain 2005].*

As a typical visual pattern recognition problem, the performance of 2D face recognition system highly depends on the features that are extracted for facial image descriptions and the classifiers used to distinguish different face identities. Moreover, face detection and alignment are indispensable components for extracting comparable features.

In the past decades, researchers in the community of biometrics, pattern recognition and computer vision, dedicated to improve the performance of 2D face recognition systems. Since the first automatic 2D face recognition system proposed by Kanade [Kanade 1973], numerous 2D face recognition algorithms have been developed. We list several milestone approaches in Fig. 1.8. Based on the taxonomy proposed in [Zhao *et al.* 2003], three categories: holistic, local feature based as well as statistical based approaches are listed in that figure.

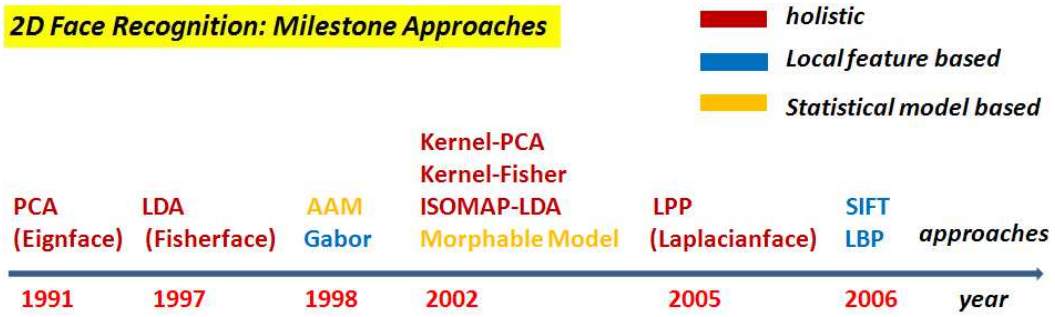


Figure 1.8: *Milestone approaches in 2D face recognition.*

The core idea of holistic approaches is the face subspace theory, which assumes that the high-dimensional pixel arrays based representation of facial images has their intrinsical low dimension structure. At the early stage, this low dimension structure is approximated by linear subspaces. Principal Component Analysis (PCA) based *Eignface* [Turk & Pentland 1991] and Linear Discriminant Analysis (LDA) based *Fisherface* [Belhumeur *et al.* 1997] are two representative approaches. At the later stage, to describe the nonlinear characteristics of face subspace, kernelization strategy is exploited, which can map the face subspace into a much higher-dimension space using nonlinear kernel functions. The nonlinear characteristics of the original face subspace is linearized in this high-dimensional space. Kernel based Principal Component Analysis (Kernel-PCA) [Yang 2002], and kernel based Fisher Discriminant Analysis (Kernel-Fisher) [Liu *et al.* 2002] are two typical approaches. Furthermore, manifold learning strategy has

also been developed, which dedicate to learn, identify and parameterize the intrinsical geometric structure of the nonlinear face manifolds. Popular approaches include the isometric mapping (ISO-MAP) [Tenenbaum *et al.* 2000] and Locality Preserving Projections (LPP) [He *et al.* 2005] *etc.*.

Different from holistic approaches, local feature based ones focus on precisely depicting of local facial regions, and compactly represent of whole facial image. Face image is often represented as some kinds of histogram-based feature vector. Face recognition approaches based on Gabor filter [Wiskott *et al.* 1997a], Scale-Invariant Feature Transform (SIFT) [Bicego *et al.* 2006], and Local Binary Patterns (LBP) [Ahonen *et al.* 2006] are the most popular ones.

Another powerful 2D face recognition approaches are based on statistical model. The main idea of these approaches comes from the theory of statical analysis of shapes. Two most famous approaches have been proposed. i) *Active Appearance Model (AAM)* [Edwards *et al.* 1998], which adopt a statistical method to learn the ways in which the shape and texture of the face vary across a set of training face images. The leaned model is parameterized by several particular sets of parameters involving model, pose, texture, expression and so on. For any test facial image, the AMM fitting process seeks the optimal set of model parameters which can be used to best generate it. Face verification or identification then can be performed by computing the distance between two sets of fitted model parameters. ii) *Morphable Models of Faces* [Blanz & Vetter 2003]: this method first constructs a 3D morphable face model by computing the correspondences of a set of 3D laser scan of faces with a reference 3D face laser scan. Similar to AAM for face recognition, given a test facial image, a fitting process is performed aiming to find a set of best model parameters, which can be used to synthesize a new face image as close as to the input one. Finally, the fitting parameters are used to perform face recognition.

1.1.4 Face Recognition: Challenges in 2D

Over the past three decades, numerous approaches for 2D face recognition have been proposed, and their performances in controlled environment have achieved a satisfactory level. However, 2D face recognition in an unconstrained daily life environment without the user's cooperation is currently a far from solved problem and a very challenging task [Li & Jain 2005]. There are still many challenging issues (e.g. expression, illumination, pose) to address [Phillips *et al.* 2000] [Zhao *et al.* 2003]. In this section, we summarize some main challenges.

Large intrasubject Variations in Facial Appearance. The intrasubject variations of facial appearance can be influenced by a multiplicity of different factors, such as facial expression, head pose, facial occlusion, illumination variations, as well as facial cosmetics *etc.*

Facial expressions are the facial changes in response to a person's internal emotion states, intentions, or social communications [Li & Jain 2005]. Intuitively, facial expression variations can change both the geometry positions of some local facial features (eyes, lips, etc.), and the topology structure of the whole facial appearance. Some impressive facial expressions examples of the previous US president *George W. Bush* are illustrated in Fig. 1.9.

Head pose has three degrees of freedom: roll, yaw, and pitch, which can be either in-

Figure 1.9: *Expressive facial expression variations of George W. Bush.*

plane or out-plane (see Fig. 1.10). In-plane rotation (i.e. roll) is a pure 2D problem which can be solved quite easily. While it will be a very challenging problem when large out-plane rotations (i.e. pitch or yaw) happen. This is mainly because that some important facial features (e.g. eye corners) might be invisible in this cases. Using multiple multi-view facial images for training might be a feasible strategy to handle this problem [Zhang *et al.* 2012].

Figure 1.10: *Comprehensive examples of head pose variations of Obama.*

Facial occlusions are usually caused by a person's facial hairs, facial accessories, hands or any other external objects (see Fig. 1.11). Since only partial face images are available, it is not surprising that the performances of face recognition systems will degrade significantly when someone displays large facial occlusions.

Facial illumination is another most challenging issue in 2D face recognition. Variations of facial illumination can largely change the intensity values of facial images. As illustrated

Figure 1.11: *Examples of the facial occlusion variations of Gaddafi.*

in each column of Fig. 1.12, the same face with the same pose looks totally different under different illuminations. These differences of appearance can be more extensive in combined with head pose variations, as shown in each row of Fig. 1.12.

Figure 1.12: *The same face looks totally different under different illuminations: 14 facial illumination variations (columns) under 9 poses (raws) (from the Yale Face Database).*

Facial cosmetics, as a common face altering technique, also have clearly impact on the performances of gray scale image-based face recognition systems [Dantcheva *et al.* 2012]. In particular, the impact due to the application of eye make up is proved to be the most pronounced. Figure 1.13 gives two examples of facial cosmetics.

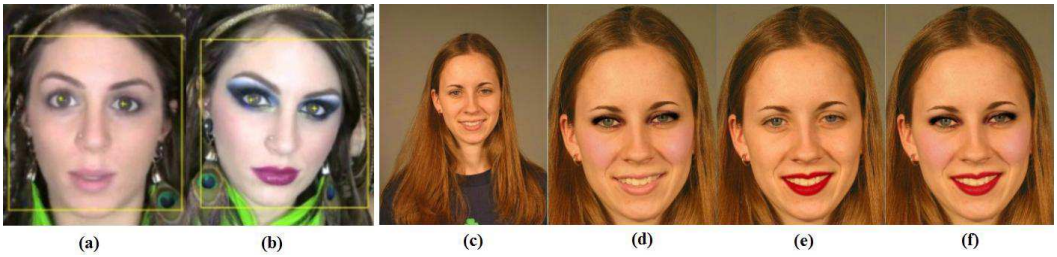


Figure 1.13: *Example of a subject in the YouTube Makeup (YMU) database, depicts the subject without [(a)] and with heavy makeup [(b)]; and example of a subject in the Virtual Makeup (VMU) database, (c) depicts the subject without makeup, whereas, (d), (e) and (f) constitute the makeup shots. The makeup shots include the synthetic addition of (d) eye makeup (e) lipstick, and (f) full makeup, respectively, using the popular Taaz software (this figure is assembled from [Dantcheva et al. 2012])*

In addition to these variations, various imaging parameters such as exposure time, aperture and lens aberrations also increase intrasubject variations.

Small intersubject Variations in Facial Appearance. Besides the large intrasubject variations, possible small intersubject variations such as identical twins (see Figure 1.14) further make 2D face recognition to be more difficult.

Figure 1.14: *From left to right, similarity of faces between twins, Obama and his imitator, a father and his son.*

Figure 1.15: *From left to right: the photos of Dzhokhar Tsarnaev's the drivers license, student ID card, and images published by the FBI and Massachusetts law enforcement during the manhunt for him and his brother.*

Moreover, video further complicates face recognition. In the scene of real video surveillance, it is a usual case that cameras tend to record low-resolution facial images. Meanwhile, the facial images may also be blurred due to the fast movement of a person's face. Figure 1.15 shows a real example of these cases. From this figure, we can find that the quality of Dzhokhar Tsarnaev's facial image (the third one) published by the FBI is very bad, which makes that it is very difficult to successfully match the photos of his drivers license and student ID card. This may be one of the most key factor which makes the face recognition technique failed in the Boston bombing manhunt.

Finally, all possible combinations of the above factors may occur at the same time in real applications, which lead to intersubject similarity smaller than intrasuject similarity in a large probability. Take the identical twins with facial expression variations as an example, the similarity between the images of two smiling twins in Figure 1.14 has a large chance to be smaller than the similarity between one face image of the smiling twins and her neutral face image.

1.1.5 Face Recognition: Opportunities in 3D

Facing so many technical challenges in 2D face recognition, one may suggest that "Why not use 3D data for face recognition?" This suggestion naturally triggers a question: "Why 3D face recognition?" To answer this question, we list several main reasons as follows.

1. *The invention and popularization of 3D scanning techniques:* With the invention and rapid development of 3D scanning technologies, the world is experiencing a 3D digital revolution. Measurement, reconstruction, scanning and reproduction of real-world 3D objects are more and more easy and popular. Nowadays, many different kinds of 3D sensors are available to capture static, even dynamic 3D data of facial surface. Please refer to next section "3D Face Acquisition Techniques" for the detailed examples.

2. *Large potential to combat 2D challenges:* Shape information based 3D face recognition approaches are naturally less susceptible to lighting conditions comparing with the photometry information based 2D face recognition approaches. With the free-form 3D facial surface in hand, it is more convenient to perform pose correction and solve the out-plane rotation problem in 2D face recognition. Moreover, facial cosmetics tend to have less impact for 3D face scans [Faltemier 2007].

3. *Facial shape and photometric appearance contain comparable and complementary information:* As pointed in [Sinha *et al.* 2006], "Many study results provide evidence that the "photometric" cues and shape cues are used about equally by humans to recognize faces". "Meanwhile, investigating the use of these cues for the recognition of familiar faces is also finding that both shape and pigmentation are about equally important ". Here the "pigmentation" means all surface reflect properties, including albedo, hue, specularity, translucency, and spatial variation in these properties, which has similar meaning as photometric cues. Moreover, as stated in [Bowyer *et al.* 2006], "we are particularly interested in 3D face recognition because it is commonly thought that the use of 3D sensing has the potential for greater recognition accuracy than 2D". Another paper states: "Depth and curvature features have several advantages over traditional intensity-based features. Specifically, curvature descriptors: (1) have the potential for higher accuracy in representing surface-based events, (2) are better suited to describe properties of the face in the area such as the cheeks, forehead, and chin, and (3) are viewpoint invariant" [Gordon 1992]; On the other hand, it has been proved many times that the performances of multi-modal 3D + 2D face recognition systems are always to be better than the one of 2D alone [Bowyer *et al.* 2006] [Mian *et al.* 2007]. That is to say, facial shape and photometric are also complementary to each other for face recognition.

1.2 3D Face Recognition

1.2.1 3D Face Acquisition Techniques

In the past decade, the techniques for scanning, capturing, reconstruction of 3D faces have witness tremendous development. In this section, we summarize three typical types of techniques: structured light, multi-view stereo, photometric stereo.

1. *Structured light:* As a "active" scanning technique, structured light has been widely used in 3D face data acquisition [Beumier & Acheroy 1999] [Huang *et al.* 2003] [Zhang & Huang 2004] [Tsalakanidou *et al.* 2005] [Zhang & Yau 2006]. The basic setup of this technique includes a camera and a light projector under given geometric relationships. The projector projects one or more encoded light patterns onto the scene. And the shape cues are extracted by measuring the deformation of light patterns on the objects' surface. Figure 1.16 illustrates an example of an encoded light patterns projected onto a face. Current structured light based 3D scanner can capture static or sequences of 3D face scans as well as 2D face images in real-time, or even with high-speed. While the drawback is that users' faces should been restricted into a limited amount of area not only covered by the projected structured light but also visible by the camera. Some popular 3D scanners based on structured light techniques are the Minolta Vivid 900/910 series [Min 2011], Inspeck Mega Capturor II 3D [Ins 2011], and Kinect camera [Kin 2011] *etc.*

Figure 1.16: *An example of structured light. A structured light (left), and its projection on to a face. Figure is collected from [Tsalakanidou et al. 2005]*

2. *Multi-view stereo:* As a "passive" capturing technique, multi-view stereo has also been widely used in 3D face data acquisition [Seitz *et al.* 2006]. This technique utilizes multiple calibrated cameras placed at various viewpoints from the subject. Under various constraints, the corresponding points of different sense images can be established and thereby the facial shape can be reconstructed. This technique does not need any flashing light, since all the camera can record the same sense at the same time with constant lights. However, accurate reconstruction of smooth surfaces (e.g. some face regions without much nature texture) is very difficult by using this technique [Bowyer *et al.* 2006]. In addition, 3D reconstruction should be operated offline due to the huge computational complexity involving this technique. Popular commercial systems which employ multi-view stereo technique are the DI3D (Dimensional Imaging) dynamic face capturing system [DI4 2011], and the 3DMD dynamic 3D stereo system[3DM 2011]. Both systems have been used for acquisition of 3D face databases [Yin *et al.* 2008b] [Benedikt *et al.* 2010] [Cosker *et al.* 2011]. Recently, Beeler *et al.* [Beeler *et al.* 2010] develop a promising 3D face capturing system based on multi-view stereo technique. This system can reconstruct facial surface with a sub-millimeter accuracy, which is comparable with those state-of-the-art systems based on structured light [Beeler *et al.* 2010]. The basic setup and some reconstructed faces by this system are depicted in Fig. 1.17.

Figure 1.17: *Left: Face model captured using a seven camera studio setup; Center: capture systems; Right: Face model captured using consumer binocular stereo camera. Figure is collected from [Beeler et al. 2010]*

3. *Photometric stereo:* Photometric stereo, originally introduced by Woodham in 1980 [Woodham 1980], is a technique for estimating the surface normals of objects by capturing a set of images of the object under different lighting conditions. Figure 1.18 shows an example of 3D facial surface reconstruction by four lights photometric stereo. The advantage of photometric stereo is that it can be used for high-speed 3D data capture. For example, the photometric stereo 3D data capture system developed by Hansen *et.al* is capable of acquiring four raw images in approximately 20 ms. However, this technique is very sensitive to the presence of projected shadows, highlights and non-uniform lighting. In addition, instead of obtaining point clouds or surface mesh directly, photometric stereo technique estimates only the surface normals of object, the following integration procedure for surface reconstruction not only adds more computation time but also introduces additional errors [Frankot *et al.* 1988]. A very recently developed photometric stereo system used for 3D face data acquisition is introduced in [Zafeiriou *et al.* 2011].

Figure 1.18: *Illustration of the photometric stereo method. From left to right: raw image set captured under four lights; estimated field of surface normals (channels x, y, z respectively); recovered depth map by performing integration on surface normals. Images are collected from [Hansen et al. 2010]*

1.2.2 Basic Concepts and Terminology

Since 3D face scan is also a biometric trait, all the concepts and terminology of general biometric recognition system are suitable to 3D face recognition except a few cases. To be more specific, we give a brief summary in this section. The general "3D face recognition" refers to two categories of application scenarios. One scenario is called "3D face

recognition" or "3D face identification". The other one is called "3D face verification" or "authentication". 3D face recognition scenario is a one-to-many matching problem while the 3D face verification scenario is a one-to-one matching problem. The rank-one recognition score and CMC curve are commonly used to evaluate 3D face recognition scenario. And the ROC curve is commonly used to evaluate 3D face verification scenario. In either scenario, a set of 3D face scans of known subjects are initially enrolled. This set is usually called the "galley" set. When performing recognition, 3D face scans of the subjects belongs to the gallery set or other subjects are called "probes", which are used to match against scans in the gallery set. 3D face recognition scenario is more challenging than the verification scenario [Bowyer *et al.* 2006]. The main reasons are listed as follows : (i) In recognition scenario, a larger gallery set tends to introduce more chances for incorrect recognition; (ii) The whole gallery set should be searched in some manner on each recognition attempt; (iii) In verification type applications, subjects are assumed to be well cooperative, while in recognition type applications, subjects may be captured without any cooperation and under uncontrolled environments. *Since this thesis focuses on 3D face recognition in the real, 3D face recognition systems proposed in this thesis are evaluated only in the 3D face recognition scenario.*

Shape and reflectance are two intrinsic properties of a face object, fortunately, current 3D face acquisition system can generally deliver 3D face scans (shape) together with their one-to-one corresponding texture counterparts (reflectance). In this case, a 2D image, also called "2D texture image" can be regarded as a "3D texture map" overlaid on a 3D face scan. A 3D face scan together with its texture map is usually called "textured 3D face". A 3D face scan can be rendered in the form of range image, point clouds or triangular mesh. A "range image", also called "depth image", is an image in which the pixel value records the distance from the sensor to the imaged facial surface. A "point clouds" records the geometry shape of a 3D face model by 3D coordinates. A "range image mesh" records the geometry shape of a 3D face model by both 3D points and their topology structures. Figure 1.19 illustrates an example of all these four kinds of face recording formats.

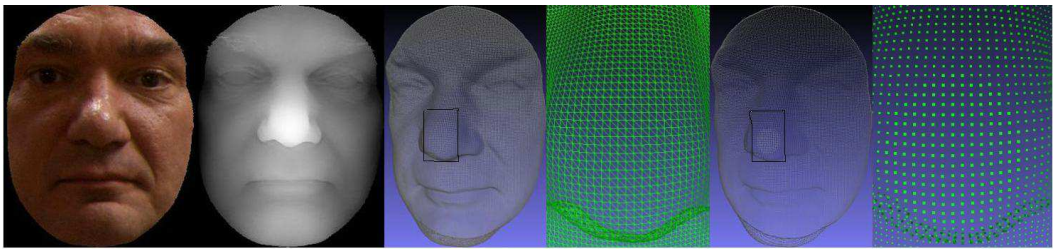


Figure 1.19: *Four kinds of 3D face recording formats, and from left to right: texture 3D face face, range image, triangle mesh and the zooming nose region, point clouds and the zooming nose region.*

According to the face modality adopted to be the gallery and probe, 3D face recognition can be performed in three matching scenarios: solely 3D shape based face recognition (3D shape vs. 3D shape), multi-modal 2D + 3D face recognition (textured 3D vs. textured 3D), and asymmetric 3D face recognition (3D shape vs. 2D texture or vice versa). Multi-modal 2D + 3D face recognition has proved to be more accurate than solely 3D shape based face

recognition [Bowyer *et al.* 2006]. *In this thesis, we focus on the matching scenario of solely 3D shape based face recognition. If there is no confusion, 3D face recognition talked in this thesis denotes the solely 3D shape based face recognition.*

As illustrated in Fig. 1.20, a typical (static) 3D face recognition system includes four steps: 3D face preprocessing (i.e. normalization), 3D face alignment (i.e. registration), 3D face feature extraction as well as 3D face matching. In particular, give a raw probe (or gallery) face scan, *3D face preprocessing* procedure contains: noise and spike removing, holes filling, nose tip localization and face cropping. As pointed out by Bowyer *et al.* [Bowyer *et al.* 2006], even under ideal illuminations for a given 3D face scanner, it is common for artifacts to occur in face regions such as oily regions (e.g. the eyes), and regions of facial hair (e.g. eyebrows, mustache, or beard). The most common artifacts are generally called "holes" or "spikes". A "hole" is essentially an facial area of missing data. A "spike" is an outlier error in the data. Holes are typically filled by interpolating new values based on the values near the boundaries of the holes. Noise and spike are usually removed by performing surface-based filters. With a relative smooth facial surface in hand, one can perform nose tip detection based on facial shape priors. Once the nose tip is available, face region can be easily cropped, and the face normalization step is finished. *3D face alignment* aims to align all the normalized face scans to a reference face scan or in a common 3D coordinate system. Iterative Closest Point (ICP) [Zhang 1994] is one of the most popular algorithm for 3D face alignment. Then, the normalized and aligned face scans are used for *3D face feature extraction*. Facial curves, local facial regions, or curvature-based facial descriptors are some commonly used facial features. Finally, for *3D face matching*, the similarity between each pair of gallery and probe scans is computed under a given distance measurement for final decision.

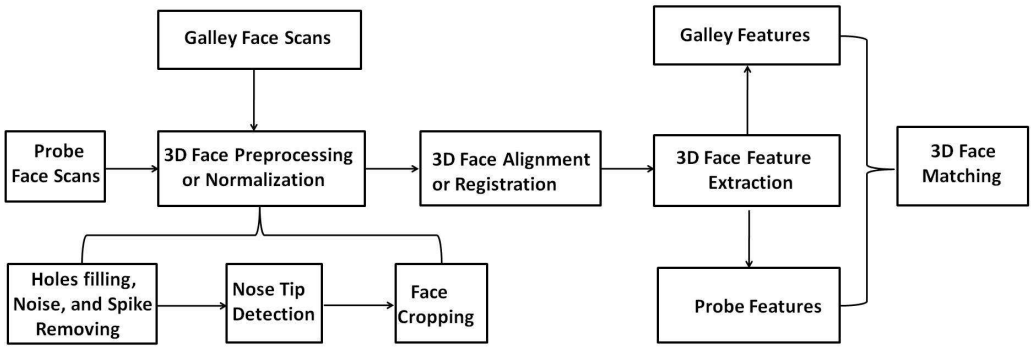


Figure 1.20: *3D face recognition processing flow.*

1.2.3 3D Face Databases

Over the past decade, many 3D face databases have been collected and released for 3D face recognition, for example, FRGC v1.0, FRGC v2.0 [Phillips *et al.* 2005], BU-3DFE [Yin *et al.* 2006a], Bosphorus [Savran *et al.* 2008], 3D-TEC [Vijayan *et al.* 2011a], UMB-DB [Colombo *et al.* 2011b], GavabDB [Moreno & Sanchez 2004], SHREC'11 [Veltkamp *et al.* 2011], UND-DB [UND 2008], etc. Please refer to Table 1.1 for a brief introduction of some public 3D face databases.

Table 1.1: *A brief introduction of some public 3D databases.*

Database	Device	Sub. Num.	Num./Sub.	Changes
FRGC v1.0	Minolta Vivid 910	275	Total: 943	-
FRGC v2.0	Minolta Vivid 910	466	Total: 4007	E
Bosohorus	Inspeck Mega Capturor II	105	Total: 4666	E, P, O
BU-3DFE	3DMD	100	4	E
3D-TEC	Minolta Vivid 910	214	2	E, Twins
SHREC'11	Roland/Escan scanner	130	Total: 780	P
GavabDB	Minolta Vivid 700	61	9	P, E
UBM-DB	Minolta Vivid 900	143	Total: 1473	E, O
UND45-J2	Minolta Vivid 910	415	Total: 1800	P
UND-F	Minolta Vivid 910	302	Total: 942	P
UND-G	Minolta Vivid 910	235	Total: 738	P

Variation labels: (P) Pose; (E) Expression; (O) Occlusion.

From Table 1.1, we can see that each 3D face database was designed for one or more specific research challenges, including:

- Models of a great number of subjects: FRGC v2.0;
- Models of stronge shape similarity: 3D-TEC;
- Models with varying facial expressions: Bosphorus, FRGCv2.0, BU-3DFE, etc;
- Models with large pose various: Bosohorus, SHREC'11, UND-J2, ect;
- Models with external occlussions: Bosohorus, UMB-DB, etc;

In this thesis, five databases are used for experimental evaluations and analysis of our proposed approaches, i.e. FRGC v1.0, FRGC v2.0, Bosohorus, BU-3DFE, and 3DTEC. Their detailed introduction can be found in chapter 3 and chapter 4.

1.2.4 Main Challenges and Objectives

In the literature, 3D face recognition is widely expected to provide solutions to overcome the unsolved issues in 2D face recognition in unconstrained environments, namely lighting variations and pose changes. However, the overwhelming majority of research works so far in the literature has targeted 3D face recognition using expressive but frontal 3D face scans (*e.g.* FRGCv2.0 database), although 3D face scans captured in an unconstrained environment can depict not only facial expression variations but also arbitrary pose changes which can lead to self-occlusion and thereby large missing face data. Moreover, they can also be severely altered by external occlusions, *e.g.* hand, scarf, hair, glasses, *etc.* These challenges are summarized as follows:

1. Facial Shape Similarities. All human faces are remarkably similar to each other, especially in terms of configuration. Therefore, the first challenge in 3D face recognition is to design effective facial representations which can capture sufficient shape detail differences between people, and thereby offer facial features discriminating enough to enable the distinction of similar 3D face scans, even in the extreme case of identical twins. Figure 1.21 illustrates an example of strong facial shape similarity by a pair of identical twins.

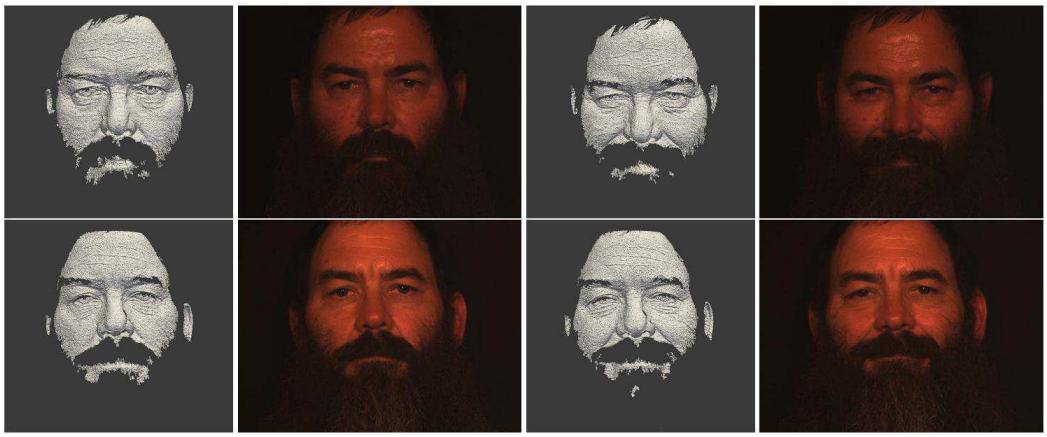


Figure 1.21: *The images of two twins taken in a session: the top row shows the first twin and the bottom row, the second; the texture images are brightened to increase visibility in this figure.*

2. Facial Expression Variations. As 3D face data accurately capture the shape information of facial surfaces, face recognition in 3D is likely to be more sensitive to facial expression variations. In real-life applications, there exist an uncountable number of facial expressions which elastically deform a facial surface, in response to a person's internal mental state, intention, or need for social communications. Such deformations of facial surfaces not only occur in non-rigid regions (e.g. mouth, see also Fig. 1.22), but also can take place in regions which are mostly considered as rigid ones (*e.g.* nose). Figure 4.26 gives a group of 3D face scans with nose region shape deformations, which are called "nose expressions" in this thesis. As arbitrary facial expressions can occur in unconstrained conditions, the second challenge of 3D face recognition in such a case is to design effective matching scheme of facial surfaces which is robust enough to facial shape deformations due to facial expressions, whether subtle, prototypical or exaggerated.

Figure 1.22: *Illustration the examples of exaggerated facial expressions with open mouth. Images taken from the Bosphrous 3D face database.*

3. Pose Changes. While the shape of a 3D object is pose invariant, its capture by existing 3D scanning technologies is pose dependent because 3D scanners only have limited field of view. As a result, a 3D face captured in an arbitrary pose can lead to self-occlusion and result in large missing data. The resultant 3D face scans are thus only partial faces.

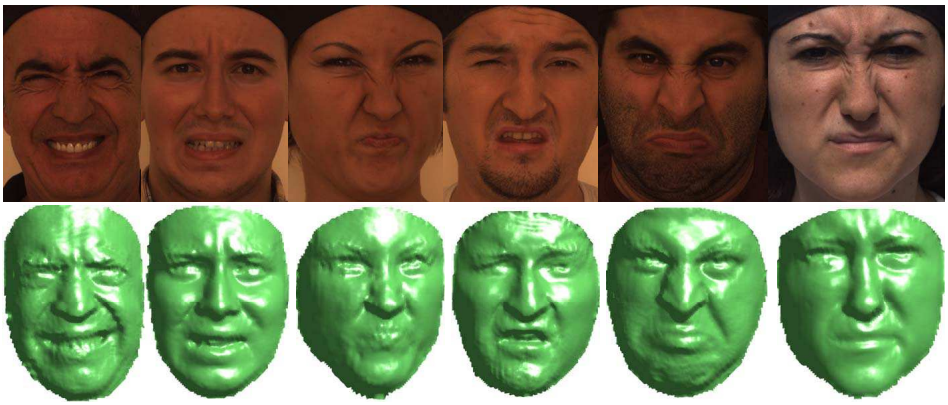


Figure 1.23: *Examples of "nose expressions" sampled from the Bosphorous database.*

In unconstrained conditions, a scanned 3D face can lie in an arbitrary pose and thereby may involve large missing data. Such partial faces make problematic the whole processing pipeline mostly designed only for frontal upright 3D face scans. For instance, 3D face matching mostly requires more or less an accurate localization of the nose tip for cropping the face region and normalizing its pose. With partial faces, this becomes problematic when the nose tip is simply invisible. Furthermore, although the initial position of a partial face is manually corrected, the ICP-based face registration algorithm still won't correctly converge due to the limited overlap region. A straight-forward example is depicted in Fig. 1.24.

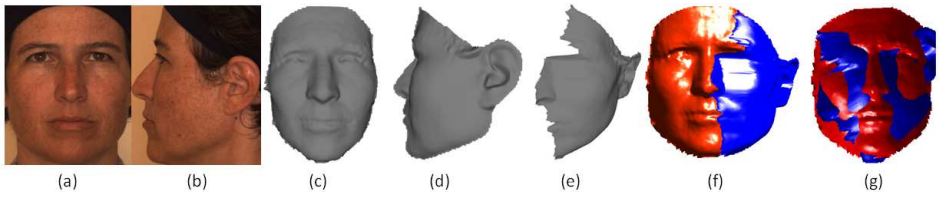


Figure 1.24: *An failed example of pose normalization: ICP registration algorithm is performed on a frontal and a partial face with* $+90^\circ$ *yaw rotation. (a,c): a frontal face scan; (b,d): left profile scan (e): manually pose normalized frontal view of (d); (f): initial position between (c) and (e); (g): (c) and (e) are not correctly registered by the ICP algorithm.*

4. External Occlusions. In unconstrained conditions, subjects are uncooperative and their face can be intentionally or covered by external objects, *e.g.*, hands, hairs, scarves, hats, *etc.* As a result, 3D face scans captured in such conditions can involve large face regions covered by external objects [Colombo *et al.* 2011a] and the use of standard 3D matching algorithms would generate high false negative rate, *i.e.*, failing to recognize the identity of "hair occluded" subjects. Figure 1.25 shows a set of 3D facial scans with various types of facial occlusions collected from the University of Milano Bicocca (UMB) 3D face database.

5. 3D Face Data of Variable Quality. 3D face scans as delivered by current 3D imaging devices are generally noisy, especially when they are captured in unconstrained conditions, and involve artifacts such as holes or spikes. Furthermore, they can vary in resolution

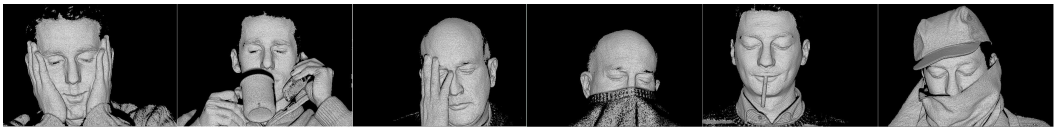


Figure 1.25: *Various facial occlusions sampled from the UMB database: e.g. hands, scarves, hats, and other types of occlusion which can occur in real-world scenarios.*

and in precision. These variations raise a very fundamental problem, *i.e.*, the stability of all these basic geometrical measurements and descriptors mostly derived from differential geometry, *e.g.*, curvatures, geodesic distance, *etc.*, which are used as building blocks in most 3D face recognition algorithms.

In this thesis, we are interested by the issue of 3D scan resolution, *i.e.*, sampling density along with triangle regularity, as the depth accuracy is related to the underlying 3D sensor and is out of the scope of this thesis. *In this thesis, when we say that a facial triangle mesh is of "low-quality" when either it has a sparse set of sampling points or it has a large set of triangles with irregular shapes, or both.* As a straightforward example, a low-quality facial triangle mesh is shown in Fig. 4.25. There are only about 5,000 points on the mesh and there are many irregular triangles (i.e. "skinny" triangles) on the facial mesh. The challenge here is the generation of 3D face meshes of good quality with theoretic guarantee of discrete measurements of differential geometry. In this thesis, we are concerned with

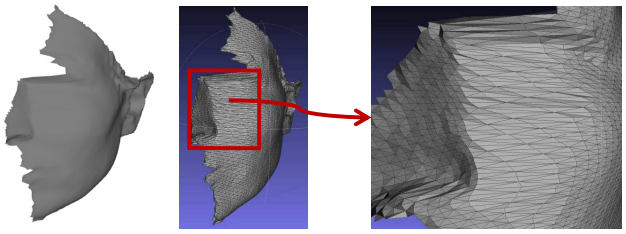


Figure 1.26: *Illustration of a low-quality data: (a): a pose normalized partial face scan; (b): triangle mesh of (a); (c): the zooming of the region in the red box of (b). We can see that the sampling density of this mesh is very low, and there are many "skinny" triangles on the mesh.*

face recognition using 3D data in unconstrained conditions. Our goal is to design 3D face recognition algorithms fine enough to be able to distinguish people with strong extra-class similarities whereas robust enough to deal with the challenges of expressions, pose and occlusions.

1.3 Thesis Contributions

1.3.1 Methodologies

3D face data accurately capture the geometric shape of facial surfaces. Therefore, a natural choice for the description of facial shape which proves to be pose invariant is to rely on the powerful tool of discrete differential geometry which provides surface geometric or

differential quantities, *e.g.*, geodesic distance between two facial points, area of a facial region, normal vector field of a facial surface, curvature values of a facial surface, or any types of scalar field function defined on a facial surface in terms of a specific discrete differential geometric operator.

However, face recognition is typically a pattern classification problem. In this work, we have extensively made use of sparse representation-based classifier (SRC) to leverage some useful prior knowledge for the purpose of face recognition.

Throughout this thesis work, local approaches have been favored with respect to holistic approaches. This deliberate choice was first motivated by our belief that the discrimination of people with similar facial surfaces, *e.g.*, identical twins, should rely on the difference of local facial details, but also by the fact that local approaches are generally more robust to those challenges of facial expressions, pose changes and occlusions in comparison with holistic approaches.

1.3.2 Main Contributions

In the following, we summarize the main contributions of this thesis work as they will be further developed in Chapters 3, 4 and 5.

a. Measuring and Learning Facial Shape Deformations for 3D Face Recognition Robust to Facial Expressions

We first developed an expression-robust 3D face recognition system which will be the focus of Chapter 3. In this work, we assume that different part of facial regions have different kinds of shape deformations for a particular expression. Based on this assumption, we believe that if the deforming behavior of each expression can be well measured and learned, then it will be very useful to improve the robustness of a 3D face recognition system to resist expression variations. To explore this insight, we design a strategy to learn the quantitative weights of different facial physical components over a training set including various expressive faces.

Specifically, the contributions of the proposed approach to dealing with facial expressions are as follows:

- We introduce a new 3D facial representation based on Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP) for facial shape description. MSMC-LNP describes the micro-structure of facial normal information in multiple scales and multiple normal component channels. The extensive experiments show that the proposed LNP based facial representation is more discriminative than both the raw normal information and the encoded range image, *i.e.*, Local Shape binary Patterns (LSP). We also show that the fusion of both multiple scales and multiple components is a helpful way to improve the final performance and demonstrate its competency on 3D face identification as well as the challenging issue of recognizing expressive identical twins.
- A learning-based strategy is proposed to find out the quantification weights of local patches of 3D facial surfaces in terms of discriminating power. Given a training database, the patch weights associated with different facial regions, encoding scales, and normal components are learned by normalizing the patch matching scores. Those scores are computed by running the sparse representation-based classifier (SRC) over MSMC-LNP features of local patches. The experimental results show that the weights associated with these local patches in 3D are quite different from those of their 2D counterparts, especially

in the nose region, thereby highlight the fact that 3D and 2D face data bear different facial information which can be further explored in a multimodal face recognition scenario.

- Using the learned patch weights along with the proposed MSMC-LNP feature, a weighted sparse representation-based classifier (W-SRC) is formulated to account for the different sensitivities of facial components to facial expressions, making the proposed 3D face recognition algorithm both discriminating enough for distinguishing identical twins and robust enough to tolerate various facial expressions. In particular, we formally establish that a weighted combination of SRCs amounts to directly weighting the corresponding feature vectors of those SRCs.
- Extensive experiments were carried out using various 3D face datasets, including FRGC v2.0, BU-3DFE, Bosphorus and 3D-TEC, to benchmark the effectiveness and the generalization skills of the proposed approach with respect to 3D face scans captured under different scenarios and conditions with different 3D sensors, depicting in particular different challenges in terms of facial expressions. This is in clear contrast with the overwhelming majority of 3D face recognition algorithms so far proposed in the literature which are only evaluated on FRGC v2.0 with 3D face scans depicting a limited number of facial expressions. The experimental results demonstrate the effectiveness and the generalization skills of the proposed approach which consistently displays competitive rank-one recognition rates over those datasets of different nature.

b. Local Geometric Feature-based Matching for 3D Face Recognition in Unconstrained Conditions

As the overwhelming majority of existing 3D face recognition algorithms, the previous work only deals with frontal expressive 3D face scans whereas 3D face data are widely expected to provide solutions to overcome those unsolved challenges in 2D face recognition, namely lighting variations and pose changes. As we have seen in the previous section, in real-life applications in unconstrained conditions, the variations of facial expressions can occur concomitantly with other challenges, *i.e.*, pose changes and external occlusions. To combat all these challenges of expression, pose and occlusion simultaneously, we propose, as extended in chapter 4, a novel local geometric feature-based matching scheme for a robust 3D face recognition algorithm. It first automatically detects a set of local salient regions spreading through a whole facial surface, then extracts a set of region-based local geometric features. The similarity between two facial surfaces is then computed by matching two sets of local geometric features. Using a preliminary version of this work, we took part to the Shrec 2011 contest for 3D face recognition and retrieval. 2 runs submitted by our group were ranked the first and the second best performance in terms of rank one recognition rate out of 14 runs submitted by four research groups.

Specifically, the contributions of this work can be summarized as follows:

- We propose a group of novel pose-invariant local shape descriptors, namely, Histograms of mesh Gradient (HOG, 1^{st} order), Histogram of Shape index (HOS, 2^{nd} order) and Histogram of Gradient of Shape index (HOGS, 3^{rd} order), and their feature-level fusion, Histogram of Multiple-order surface differential Quantities (HOMQ).
- We provide a group of compact facial surface representations. Each facial surface can be represented as a sparse set of automatically detected salient vertices associated with the proposed local shape descriptors. They are Multiple Vertex based descriptions of multiple order differential geometry quantities, namely gradient, shape index and shape index derivative, denoted as MV-HOG (1^{st} order), MV-HOS (2^{nd} order), MV-HOGS (3^{rd} order),

and MV-HOMQ (multi-order), respectively.

- A fine grained local feature matcher (FGM) using the Sparse Representation based Classifier (SRC) [Wright *et al.* 2009a] is introduced and compared with a coarse grained local feature matcher (CGM) that was used in our preliminary work [Li *et al.* 2011c]. CGM performs the matching of two face scans through their corresponding sparse sets of local shape descriptors and simply counts the number of matched vertex pairs as matching score. This is problematic when the underlying face scans have severe missing data due to pose and/or external occlusions, which result in a significant decrease of the number of matched vertex pairs. In contrast, FGM alleviates this vulnerability of missing data by making use of SRC to compute a mean reconstruction error when each local shape descriptor from a probe face scan is sparsely reconstructed from an over-complete dictionary composed of the local shape descriptors extracted from all the face scans in the gallery.
- All the previous steps are directly carried out on mesh-based 3D face scans without any prior registration or learning. The overall result is a uniform registration free 3D face recognition framework which can deal with facial expression variations, arbitrary pose changes and occlusions at the same time. This is in clear contrast to the overwhelming majority of 3D face recognition algorithms which only deal with expressive frontal face scans and mostly require in addition an expensive registration step.
- Comprehensive experiments were carried out on the Bosphorus database, the largest public database which best accounts for the various challenges of 3D face recognition in uncontrolled conditions, i.e. expression variations, pose changes and external occlusions. The experimental results demonstrate the effectiveness of the proposed method which achieves the best state-of-the-art rank-one recognition rates over the whole dataset (**96.56%**) and its three subsets of expression variations (**98.82%**), pose changes (**91.14%**) and occlusions (**99.21%**). Additional experiments on the FRGC v2.0 database show that the proposed method consistently displays competitive performance in comparison with the state-of-the-art algorithms which mostly require accurate face alignment and/or learning of discriminating features, and thereby demonstrates its generalization skill.

c. Dealing with 3D Face Data of Variable Quality

In the previous work, we made use of surface differential quantities of several orders computed over 3D mesh face data in order to characterize local facial shape variations. Their stability thus widely depends upon the quality of the underlying 3D face meshes. Unfortunately, 3D face scans captured in unconstrained conditions could be delivered by different 3D sensors and thus are typically of variable quality, *i.e.*, sparse set of sampling points and/or large set of non-uniform triangles. In this thesis work, we further studied the issue of the efficient generation of high-quality triangle meshes with theoretic guarantee of the measurements of various differential geometry quantities over discrete triangle meshes. This can be achieved using a Delaunay refinement based surface meshing algorithm belonging to one of the two following categories:

1. The sampling is computed in $\mathbb{R}^3$, and triangulated using volumetric Delaunay triangulation, such as [Amenta & Bern 1998] [Amenta *et al.* 2000] [Boissonnat & Oudot 2005].

2. The sampling is directly computed on a smooth surface, such as geodesic Delaunay refinement [Dai *et al.* 2007], and Chew's second algorithm [Paul Chew 1993].

However, for the sake of efficiency and simplicity, we rather consider Delaunay refinement based surface algorithms with the sampling computed in the conformal parameter domain, and triangulated using planar Delaunay triangulation, such as [Alliez *et al.* 2003] [Marchandise *et al.* 2012] [Alliez *et al.* 2008], which can be classified as the third category. Indeed, as compared to the previous two categories of mesh refinement approaches, the planar Delaunay refinement method is much easier to implement, and the data structure for planar triangulation is much simpler than that of tetrahedral mesh as used in the first category, and the planar algorithm is much more efficient. However, while the theoretic convergence of surface normals and curvature measures for the algorithms of the previous two categories has been thoroughly discussed in [Amenta & Bern 1998] [Boissonnat & Oudot 2005] [Cohen-Steiner & Morvan 2003a] [Cohen-Steiner & Morvan 2006] [Morvan & Thibert 2004] by Amenta, Bern, Boissonnat, Oudot, Cohen-Steiner, Morvan, and Thibert, all the corresponding theoretical results of the mesh refinement algorithms using conformal mapping remain widely open.

As a result, we also provided as the third contribution (chapter 5) the theoretic guarantee of conformal mapping-based surface meshing algorithms regarding the convergence of various differential geometry quantities, mainly through the following two theorems:

Theorem 1.3.1 (Delaunay Refinement on Constant Curvature Surfaces)**.** *Let M be a compact Riemannian surface with constant curvature. Suppose that the boundary of M is empty or is a union of geodesic circles. For any given small enough $\varepsilon > 0$, the Delaunay refinement algorithm terminates. Moreover, the shortest edge length is greater than ε.*

Remark

- As an obvious consequence of this theorem, in the final triangulation, all triangles are well-sized and well-shaped.
- This theorem gives a Delaunay refinement algorithm for abstract Riemannian surfaces with constant curvature. It can be easily extended to any compact Riemannian surfaces, by using the classical uniformization theorem. In fact, it is well known that any Riemannian metric defined on a smooth (compact with or without boundary) surface M can be conformally deformed into a metric of constant curvature $c \in \{-1, 0, 1\}$ (i.e. uniformization metric), the constant c depending on the topology of M. So, if M is any compact Riemannian surface, we first use the uniformization theorem to deform the metric to a metric with constant curvature, then we apply the Delaunay refinement algorithm described above, and finally, we apply the inverse transformation. The angles of the mesh do not change under the conformal transformation and the length of the edges are multiplied by a bounded factor.

Our next step is to compare the geometry of a smooth surface embedded in E^3, and the geometry of a well shaped and well sized triangulation (generated by a Delaunay refinement algorithm on conformal uniformization domain). The definition of the curvatures of a smooth surface M is standard. On the other hand, there are many possible choices to define different types of curvatures on a triangulation T. Here, we choose to use the *normal cycle theory*, since it gives a general framework involving both smooth and discrete approaches. This implies that we deal with *curvature measures*, in the sense of

[Cohen-Steiner & Morvan 2003a], which allow to compare the local (but not pointwise) curvatures of M and T.

To compare the curvatures of an open set U of M with the curvature of a "corresponding" open set V of T, we need a bijection between U and V. Usually, this one to one correspondence is the orthogonal projection π from V to $U = \pi(V)$, well defined if the vertices of T are dense enough in M. however, if M is defined by a parameterization, we introduce a new bijection η as follows:

Let $(M, \mathbf{g})$ be a C^2-smooth Riemannian surface. Let $\mathscr{D}$ be the unit disk on the uv-plane. Let $\varphi : \mathscr{D} \to M$ be a conformal parameterization of M. Let $\mathscr{T}$ be a triangulation of the parameter domain $\mathscr{D}$. The conformal mapping φ induces a triangulation $\varphi(\mathscr{T})$ of the smooth surface M, denoted by T. Each vertex $p_i \in \mathscr{T}$ corresponds to a vertex $\varphi(p_i) \in T$, each triangle on $\mathscr{T}$ corresponds to a face in T. This induces a piecewise linear map $\tau : \mathscr{T} \to T$.

Définition 1 (Natural projection)**.** *The homeomorphism from the triangle mesh T to the smooth surface M*

$$\eta = \varphi \circ \tau^{-1} : T \to M$$

is called the natural projection.

Our main result is the following

Theorem 1.3.2 (Main Theorem)**.** *Let M be a compact Riemannian surface embedded in $\mathbb{E}^3$ with the induced Euclidean metric, T the triangulation generated by Delaunay refinement on conformal uniformization domain, with a small enough circumradius bound ε. Suppose $q \in T$, then*

$$|\mathbf{n}(q) - \mathbf{n}(\eta(q))| \quad = O(\varepsilon) \tag{1.1}$$
$$|\mathbf{n}(q) - \mathbf{n}(\pi(q))| \quad = O(\varepsilon). \tag{1.2}$$

Suppose B is the relative interior of a union of triangles of T, then:

$$|\phi_T^G(B) - \phi_M^G(\pi(B))| \quad \leq K\varepsilon \tag{1.3}$$
$$|\phi_T^H(B) - \phi_M^H(\pi(B))| \quad \leq K\varepsilon \tag{1.4}$$
$$|\phi_T^G(B) - \phi_M^G(\eta(B))| \quad \leq K\varepsilon \tag{1.5}$$
$$|\phi_T^H(B) - \phi_M^H(\eta(B))| \quad \leq K\varepsilon \tag{1.6}$$

where for fixed M

$$K = O(\sum_{\{t \in T, t \subset \bar{B}\}} r(t)^2) + O(\sum_{\{t \in T, t \subset \bar{B}, t \cap \partial B \neq \emptyset\}} r(t)),$$

$r(t)$ being the circumradius of triangle t. Moreover, K can be further replaced by:

$$K = O(area(B)) + O(length(\partial B)).$$

Here $\mathbf{n}(\cdot)$ denotes the normal vector field, and $\phi^G(\cdot)$ and $\phi^H(\cdot)$ denote the Gaussian and mean curvature measures defined in [Cohen-Steiner & Morvan 2003a].

Remark To the best of our knowledge, this is the first theoretical result on the convergence of surface normal and curvature measures for the meshing algorithms in the third category.

1.4 Thesis Organization

The remainder of this thesis is organized as follows.

- In Chapter 2, we review the most representative and successful approaches for 3D face recognition. Furthermore, their robustness to the challenges of facial expression, pose, and occlusion variations, as well as the "low-quality data" issue are discussed.
- In Chapter 3, we introduce the proposed 3D face recognition approach to solve facial expression variations, including local normal patterns based face description, facial patch weights learning strategy, sparse representation, and experimental results on the FRGC v2.0, Bosphorus (only fontal scans), BU-3DFE and 3D-TEC databases.
- In Chapter 4, we present the proposed registration free 3D face recognition method, namely, local feature detection, description and matching framework, its performance on the Bosphrous and FRGC v2.0 databases, and its effectiveness to handle facial expressions, occlusions and pose variations.
- In Chapter 5, regarding to the "low-quality data" issue, we describe we describe the surface meshing algorithm based on planar Delaunay refinement and conformal mapping, and the theoretic results related to the convergence of surface normal and curvature measures, as well as their numerical evaluations.
- In Chapter 6, we summarize our thesis and propose future directions.

CHAPTER 2

Literature Review: 3D Face Recognition

Contents

2.1 Expression Deformation Modeling based Approaches 29
2.1.1 Thin-Plate-Spline (TPS) based Deformation Modeling 30
2.1.2 Annotated Deformable Model based Deformation Modeling 32
2.1.3 PCA Subspace based Deformation Modeling 38
2.1.4 3D Morphable Model based Deformation Modeling 40
2.1.5 Bilinear Model based Deformation Modeling 41
2.2 Surface-distance based Approaches 43
2.2.1 Iso-metric Surface based Approaches 43
2.2.2 Iso-level Curves based Approaches 45
2.2.3 Iso-geodesic Strips based Approaches 48
2.3 Local Region/Feature based Approaches 49
2.3.1 Multiple Region Matching based Approaches 50
2.3.2 Local Geometry-Texture Descriptor based Approaches 57
2.3.3 Local Geometry-Shape Descriptor based Approaches 60
2.4 Discussion 64
2.5 Summary 65

Along with the development in 3D scanning and capturing techniques, the task of recognizing 3D face scans has been discussed in many ways, leading to numerous approaches. Similar to the taxonomy of 2D face recognition, 3D face recognition approaches can also be generally classified into two categories: *i.e.* holistic based and local feature-based. In this chapter, we review the most representative and successful approaches for 3D face recognition. To be more specific, we classify the holistic based approaches into two categories: 1. expression deformation modeling based approaches; 2. surface-distance based approaches; and detail the local feature-based approaches as 3. local region/feature based approaches. Similar taxonomy can be found in [Drira *et al.* 2013].

2.1 Expression Deformation Modeling based Approaches

Expression deformation modeling based approaches generally assume that human facial surfaces are non-rigid and deformable surfaces, deforming according to the variation of facial expressions. The deformation process between faces can be simulated by

some physical principles such as the *Thin-Plate-Spline based fitting* [Lu & Jain 2008], the *Annotated Deformable Model based fitting* [Kakadiaris *et al.* 2007a] and the *deformable model based fitting* [Mpiperis *et al.* 2008a]. Meanwhile, the deformation patterns can be learned by some statistical models such as the *PCA Subspace based Deformation Modeling* [Al-Osaimi *et al.* 2009], *3D Morphable Model based Deformation Modeling* [Amberg *et al.* 2008], and *Bilinear Model based Deformation Modeling* [Mpiperis *et al.* 2008a]. Two main properties of the expression deformation modeling based approaches are summarized as follows:

- a deformable model are usually generated for rigid registration and non-rigid fitting, resulting pose normalized and vertex-number normalized (*i.e.* one-to-one correspondence between model and scan) face scans;
- a parametrical stastical model are usually built for expression morphing, transfer, and neutralization.

2.1.1 Thin-Plate-Spline (TPS) based Deformation Modeling

Lu and Jain extend their previous ideas of using an ICP for rigid face matching [Lu *et al.* 2004] and a Thin-Plate Spline (TPS) based deformation estimator for non-rigid face matching [Lu & Jain 2005], and propose an expression robust 3D face matching approach by modeling an expression deformation model using the TPS function [Lu & Jain 2006] [Lu & Jain 2008]. As shown in Figure 2.1, their method includes four basic steps: (1) Hierarchical facial surface sampling for deformation learning; (2) TPS-based deformation transfer and synthesis; (3) Expression-specific and/or expression-generic deformable models constructions; (4) Deformable model fitting for face matching. The core

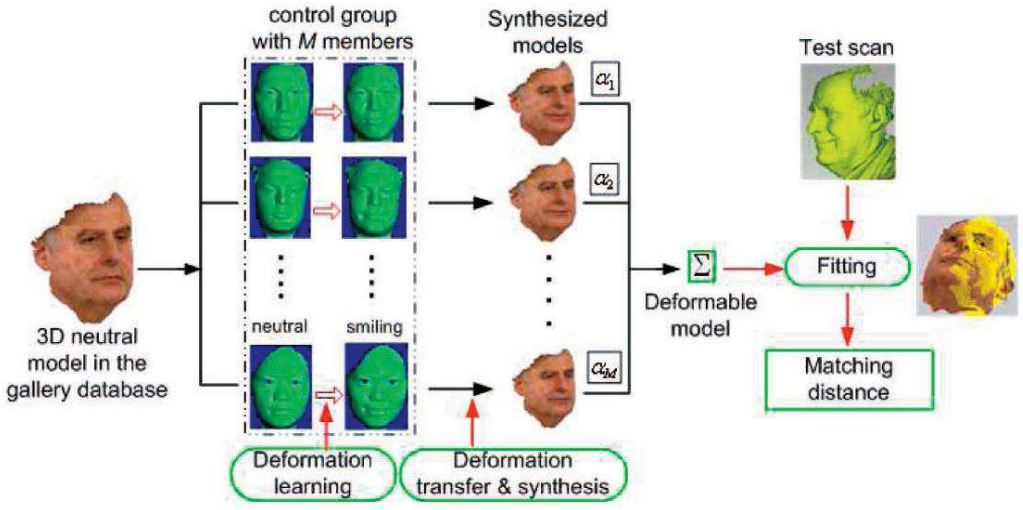


Figure 2.1: *Pipleline of the thin-plate-spline based deformation modeling approach [Lu & Jain 2008].*

part of their method is the step (2), which can be easily understood based on the following

diagram.

$$\begin{array}{ccc} S_{ne} & \xrightarrow{f} & M_{ne} \\ \downarrow \phi & & \downarrow \varphi = g \circ \phi \circ f^{-1} \\ S_{ex} & \xrightarrow{g} & M_{ex} \end{array}$$

In this diagram, suppose S_{ne} and S_{ex} are a pair of neutral and non-neutral scans of the same person from the training set (i.e. control group), and suppose M_{ne} is a 3D neutral model from the gallery set. The goal of the TPS-based deformable model is to learn the expression deformation between the pair of neutral and non-neutral scans of the person in the training set (i.e. $\phi : S_{ne} \to S_{ex}$), and transfer the learned deformation to another person in the gallery set by the mappings f and g, finally, generate new expression model of M_{ne} by the mapping φ (i.e. $\varphi = g \circ \phi \circ f^{-1} : M_{ne} \to M_{ex}$).

To find the mappings, in practice, facial surfaces S_{ne}, S_{ex} and M_{ne} are hierarchically sampled as the same numbers corresponding facial landmarks (See Fig. 2.2). Then, the following procedures are performed.

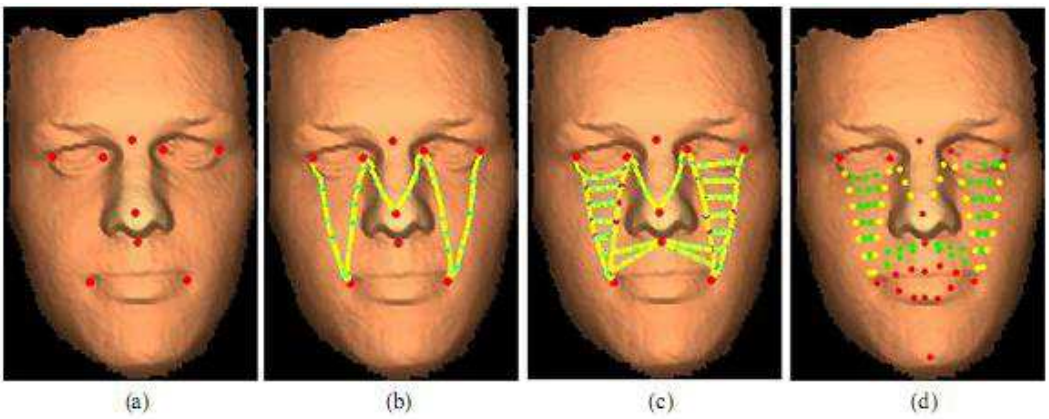


Figure 2.2: *Hierarchical surface sampling. (a) first layer (fiducial set); (b) second layer; (c) third layer; (d) final landmark set.*

(1) Register the non-neutral scan S_{ex} with the neutral scan S_{ne} and establish the mapping $\phi : S_{ne} \to S_{ex}$. ϕ is assume to be a TPS mapping, which can be estimated from the corresponding facial landmarks.

(2) Similar to ϕ, establish the mapping $f : S_{ne} \to M_{ne}$ from the landmark set of S_{ne} to that of M_{ne}.

(3) Use the same mapping $g = f$ to transfer the landmark set of S_{ex} to M_{ne} as M_{ex} .

(4) Establish the mapping $\varphi : M_{ne} \to M_{ex}$ by the composite mapping $\varphi = g \circ \phi \circ f^{-1}$.

Given two sets of corresponding landmarks $U = (u_1, u_2, \ldots, u_m)^T$ and $V = (v_1, v_2, \ldots, v_m)^T$ sampled from two surfaces, the TPS mapping F is given by the conditions

$$F(u_i) = v_i, i = 1, 2, \cdots, m.$$

Meanwhile, F can be represented as

$$F(u) = c + A \cdot u + W^T s(u),$$

where c, A and W are TPS parameters; $s(u) = (\sigma(u-u_1), \sigma(u-u_2), \cdots, \sigma(u-u_m))^T$ and $\sigma(r) = |r|$.

Suppose there are N pairs of neutral and non-neutral training scans of N subjects (i.e. $\{S_{ne}^i\}_{i=1}^N$ and $\{S_{ex}^i\}_{i=1}^N$) and one 3D neutral model M_{ne} in the gallery set. Then, by running step (2), N expression deformation mappings $\{\phi\}_{i=1}^N$, N deformation synthesis mappings $\{\varphi\}_{i=1}^N$, and N synthesized deformation models $\{M_{ex}^i\}_{i=1}^N$ can be achieved. By combing all the N synthesized models, one can construct the deformable model M of the gallery subject by

$$M = M_{ne} + \sum_{i=1}^{N} \alpha_i \cdot (M_{ex}^i - M_{ne}) \tag{2.1}$$

where α_i's are the mixing weights. If all the N training pairs have the same expression, M can be an expression-specific model; otherwise, M can be an expression-generic model. The model fitting is formulated as an optimization problem to minimize the cost function

$$E(\alpha_i, \ldots, \alpha_N; R, T) = \|M_{ne} + \sum_{i=1}^{N} \alpha_i \cdot (M_{ex}^i - M_{ne}) - \xi(S_{te}|R,T)\|^2,$$

where R and T are the rotation and translation matrices; S_{te} is a test scan and $\xi(S_{te}|R,T)$ represents applying (R,T) to S_{te}. For each iteration, rigid transformations (R,T) can be solved by the iterative closest point (ICP) algorithm and then the parameters α_i can be solved by gradient-based iterative method. Once the model fitting is finished, the root-square distance computed by the ICP algorithm is used as the matching distance.

For experimental evaluation, a small training set consisting of 10 subjects, each with 7 expressions (neutral, happy, angry, smile, surprise, deflated, and inflated), is collected in their laboratory (MSU). Meanwhile, considering the time cost of model fitting, only 877 scans of 100 subjects from the FRGCv2.0 database are selected for testing, and the experimental results show that the proposed deformation modeling scheme can resist expression variations to some extent.

As pointed out by the authors, the obvious disadvantages of their current method are i) The fiducial facial landmarks required for deformation learning are manually extracted; ii) Only a small training set (10 subjects) are used which makes the deformable model has a very limited generalization ability; iii) Since the ICP process has to be repeated during the model fitting stage, their method is quite time-consuming; iv) A limited experimental results are reported on a subset of the FRGCv2.0 database.

2.1.2 Annotated Deformable Model based Deformation Modeling

Kakadiaris *et al.* [Kakadiaris *et al.* 2007a] extend their pervious work [Passalis *et al.* 2005], and propose an annotated deformable model approach for 3D face recognition in the presence of facial expressions.

The pipeline of their method is as follows: (1) Annotated face model (AFM) building. AFM is achieved by averaging a set of 3D facial meshes. Meanwhile, using the knowledge of facial physiology, it is segmented into different anatomical areas by a set of anthropometric landmarks (Figure 2.3 (a) and (b)); (2) AFM based multi-stage, and coarse-to-fine rigid face alignment. Particularly, each preprocessed face scan is aligned with the AFM by using *spine image*, *Iterative Closest Point (ICP)*, and *Simulated Annealing on Z-Buffers* in

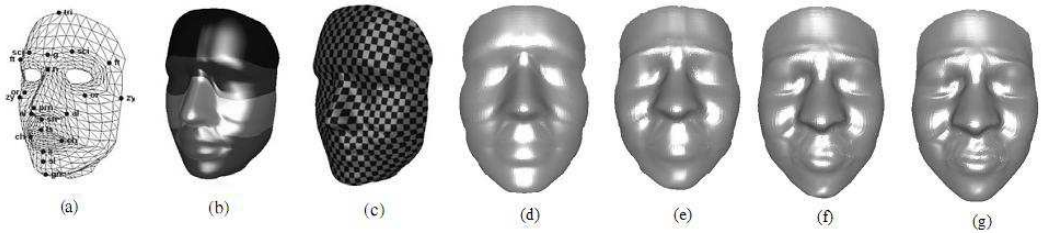


Figure 2.3: *Annotated face model and its fitting process: (a) Anthropometric landmarks for face annotation; (b) Annotated facial areas; (c) Checkerboard texture to demonstrate the global UV parameterization; AFM after (d) 0, (b) 8, (c) 32, and (d) 64 iterations. Figures are taken from [Passalis et al. 2005] and [Kakadiaris et al. 2007a].*

sequence; (3) AFM based deformable fitting. Particularly, the AFM is fitted to each aligned face scan using the *Elastically adapted Deformable Model* [Terzopoulos & Metaxas 1991] [Metaxas & Kakadiaris 2002]; (4) AFM based geometry image extraction. After the fitting process, each deformed AFM is flatted into $\mathbb{R}^2$ by a spherical mapping function to achieve a global UV parameterization. Then, a geometry image [Gu *et al.* 2002] and a normal map image are extracted from the parameterized AFM; (5) Wavelet based geometry image analysis. Haar transformation and pyramid transformation are performed on the geometry images to extract wavelet coefficients; (6) Face matching based on the Haar metric and the Pyramid metric. The core part of their method for handling facial expressions is step (3), where the Elastically adapted Deformable Model is used. To deeply understand the basic principle of the model, in our thesis, we give a brief review in the following paragraphs. For more details, we refer readers to [Terzopoulos & Metaxas 1991].

Elastically adapted Deformable Model has the following analytical formulation

$$M_q\ddot{q} + D_q\dot{q} + K_q q = g_q + f_q \tag{2.2}$$

where M_q is the mass matrix, D_q is the damping matrix, K_q is the stiffness matrix, g_q gives the centrifugal and Coriolis forces, f_q are the external forces, and q is a vector gives the model parameters. Equation 2.2 is a Lagrangian equation of motion, which governs both global and local deformation behaviors of deformable superquadrics. Following [Terzopoulos & Metaxas 1991], we summarize here the deviation of this equation.

1. Geometry of the Model

As shown in Fig. 2.4, the models are closed surfaces whose intrinsic coordinates are $u = (u,v)$, defined on a domain Ω. The positions of points on the model relative to an inertial frame Φ are given by a vector-valued, time-varying function of u:

$$x(u,t) = (x_1(u,t), x_2(u,t), x_3(u,t))^T. \tag{2.3}$$

Meanwhile, these positions relative to the model frame φ can be expressed as

$$x(u,t) = c(t) + R(t)p(u,t) \tag{2.4}$$

where $c(t)$ is the origin of φ at the center of the model, rotation matrix $R(t)$ gives the orientation of φ relative to Φ, and $p(u,t)$ denotes the position of points on the model

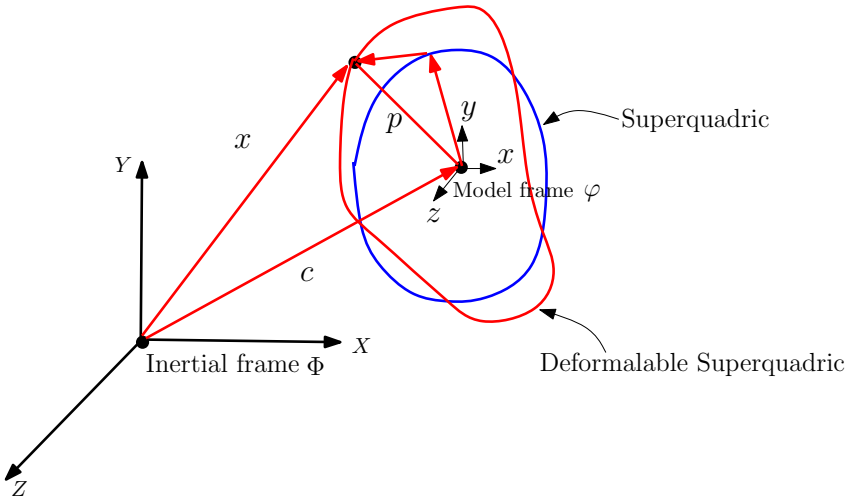


Figure 2.4: *Geometry of a deformable superquadric.*

relative to the model frame φ. Considering the shape deformations, $p(u,t)$ can be further expressed as

$$p(u,t) = s(u,t) + d(u,t) \tag{2.5}$$

where $s(u,t)$ is a reference shape and $d(u,t)$ is a displacement function.

In [Terzopoulos & Metaxas 1991], the authors employed superquadric ellipsoids as the initial reference shape,which can be expressed as

$$s = a \begin{pmatrix} a_1 C_u^{\varepsilon_1} C_v^{\varepsilon_2} \\ a_2 C_u^{\varepsilon_1} S_v^{\varepsilon_2} \\ a_3 S_u^{\varepsilon_1} \end{pmatrix}. \tag{2.6}$$

where $u \in [\pi/2, \pi/2]$ and $v \in [-\pi, \pi]$, and where $S_w^{\varepsilon} = \text{sgn}(\sin w |\sin w|^{\varepsilon})$ and $C_w^{\varepsilon} = \text{sgn}(\cos w |\cos w|^{\varepsilon})$, respectively. Here, $a \geq 0$ is a scale parameter, $0 \leq a_1, a_2, a_3 \leq 1$ are aspect ratio parameters, and $\varepsilon_1, \varepsilon_2 \geq 0$ are "squareness" parameters. All these model parameters can be collected into a parameter vector

$$q_s = (a, a_1, a_2, a_3, \varepsilon_1, \varepsilon_2)^T. \tag{2.7}$$

In general, one can express the displacement d as a linear combination of a finite number of basis functions $b_i(u)$

$$d(u,t) = \sum_{i=1}^{l} \text{diag}(b_i(u)) q_i(t). \tag{2.8}$$

Equation 2.8 can further be written as

$$d = S q_d \tag{2.9}$$

where S is the shape matrix whose entries are the basis functions and $q_d = (\cdots, q_i, \cdots)^T$ is the vector form of the generalized coordinates q_i.

2. Kinematics of the Model

The velocity of points on the model is given by

$$\begin{aligned} \dot{x} &= \dot{c} + \dot{R}p + R\dot{p} \\ &= \dot{c} + B\dot{\theta} + R\dot{s} + RS\dot{q}_d \end{aligned} \tag{2.10}$$

where $\theta = (\cdots, \theta_i, \cdots)^T$ denotes the rotational coordinates of the model, and $B = [\cdots \partial(Rp)/\partial\theta_i \cdots]$. Furthermore,

$$\dot{s} = [\frac{\partial s}{\partial q_s}]\dot{q}_s = J_s\dot{q}_s \tag{2.11}$$

where J_s is the Jacobian of S. Finally, we can write

$$\dot{s} = [I\ B\ RJ_s\ RS]\dot{q} = L\dot{q} \tag{2.12}$$

where $q = (q_c^T, q_\theta^T, q_s^T, q_d^T)^T$, with $q_c = c$ and $q_\theta = \theta$.

3. Dynamics of the Model

When fitting the model to visual data, the aim is to find the vector q, which involves the degrees of freedom of the model. The components q_c and q_θ denote the global rigid motion coordinates, q_s denote the global deformation coordinates, and q_d denote the local deformation coordinates of the model. To make the model dynamic in q, the authors use a physically based way to introduce mass, damping, and a deformation strain energy and build a set of equations of motion under the action of externally applied forces. The deviation of the motion equation (2.2) is introduced as follows.

a. Kinetic Energy: Mass Matrix. The kinetic energy of the model is given by

$$T = \frac{1}{2}\int \mu\dot{x}^T\dot{x}du = \frac{1}{2}\dot{q}^T[\int \mu L^T L du]\dot{q} = \frac{1}{2}\dot{q}^T M_q\dot{q} \tag{2.13}$$

where $M_q = \int \mu L^T L du$ is the symmetric mass matrix, and $\mu(u)$ is the mass density of the object, respectively.

b. Energy Dissipation: Damping Matrix. The velocity dependent kinetic energy dissipation can be expressed in terms of the (Raleigh) dissipation functional:

$$F = \frac{1}{2}\int \gamma\dot{x}^T\dot{x}du = \frac{1}{2}\dot{q}^T[\int \gamma L^T L du]\dot{q} = \frac{1}{2}\dot{q}^T D_q\dot{q} \tag{2.14}$$

where $\gamma(u)$ is a damping density, and D_q is the damping matrix which has the same form as M_q, except that γ replaces μ.

c. Strain Energy: Stiffness Matrix. The elastic properties of the model can be defined in terms of a deformation energy with the general form

$$E = \frac{1}{2}q^T K_q q \tag{2.15}$$

where

$$K_q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 \\ & & K_{ss} & K_{sd} \\ symmetric & & & K_{dd} \end{pmatrix}. \tag{2.16}$$

is the stiffness matrix, and the submatrices K_{ss}, K_{sd} and K_{dd} related to the global deformation parameter q_s and the local one q_d, respectively.

d. External Forces and Virtual Work. The external forces $f(u,t)$ applied to the model do virtual work, which can be written as

$$\delta W_F = \int f^T L \delta q du = \int f_q \delta q du \tag{2.17}$$

where

$$f_q = f^T L = (f_c, f_\theta, f_s, f_d) \tag{2.18}$$

with

$$\begin{array}{ll} f_c^T = \int f^T du, & f_s^T = \int f^T R J_s du \\ f_\theta^T = \int f^T B du, & f_d^T = \int f^T R S du \end{array} . \tag{2.19}$$

e. Lagrange Equations of motion. The Lagrange equations of motion can be written as

$$\frac{d}{dt}(\frac{\partial T}{\partial \dot{q}})^T - (\frac{\partial T}{\partial q})^T + (\frac{\partial F}{\partial \dot{q}})^T + \delta_x E = f_q. \tag{2.20}$$

The first two terms of (2.20) express the inertial forces and can be written as

$$\frac{d}{dt}(\frac{\partial T}{\partial \dot{q}})^T - (\frac{\partial T}{\partial q})^T = M_q \ddot{q} - g_q \tag{2.21}$$

where

$$g_q = -\dot{M}_q \dot{q} + \frac{1}{2}[\frac{\partial}{\partial q}(\dot{q}^T M_q \dot{q})]^T \tag{2.22}$$

gives the centrifugal and Coriolis forces. The third term expresses the friction forces and can be written as

$$\frac{\partial F}{\partial \dot{q}} = D_q \dot{q}. \tag{2.23}$$

And the forth term, the variational derivative of E with respect to x, expresses the elastic forces

$$\delta_x E = K_q q. \tag{2.24}$$

Substituting the above equations into (2.20) yields the equations of motion (2.2). The numerical solution of (2.2) can be approximated by using a Finite Element Method. Please refer to [Terzopoulos & Metaxas 1991] for more details of the discretization of the model.

In [Kakadiaris *et al.* 2007a], the authors report a 97.0% rank-one recognition rate and a 97.0% verification rate at 0.1% FAR on the FRGC v2.0 dataset. Along with the competitive performance, additional experimental results tested on the subsets (i.e. neutral and non-neutral) of the FRGCv2.0 database and the UH database show that their face recognition system is also robust to the facial expression variations and the face sensor changes. Furthermore, their system is very efficient in terms of both space and time. It takes 15 seconds to finish all the first five steps for a raw face scan, and only 1/1,000 seconds to compare the wavelet coefficients of two faces on a typical modern PC.

With the above superiorities, the method in [Kakadiaris *et al.* 2007a] also has some disadvantages. For example, the computation of pyramid transform is quite expensive. Meanwhile, the dimensionality of the wavelet coefficients will be very high when multiple

layers wavelet transform is used. To solve these problems, in [Ocegueda *et al.* 2011b], a Markov Random Field (MRF) model is formulated to select the most discriminative wavelet coefficients and to achieve a more compact signature. It is clear that by using the compact signatures, computational efficiency of the recognition system can be improved in terms of both storage requirements and computational time. Similar idea is used in [Ocegueda *et al.* 2011a], by using Linear Discriminant Analysis based dimensionality reduction techniques, the original high dimensional wavelet coefficients (*i.e.*, 7200) is reduced to quite low dimensionality (*i.e.*, 57). With such a compact and discriminative signature and different training databases, this improved method achieves a better performance ($\geq$ 99%) on the FRGCv2.0 database.

Recently, towards real-world 3D face recognition, annotated deformable model based recognition framework has also be extended to handle large pose variations for partial data based 3D face recognition [Perakis *et al.* 2009] [Passalis *et al.* 2011] [Chu *et al.* 2013].

Compared to their previous method [Kakadiaris *et al.* 2007a], the main contributions in [Perakis *et al.* 2009] and [Passalis *et al.* 2011] are: (1) A novel 3D landmark detector is introduced for pose estimation and registration of partial faces; (2) A symmetric deformable model fitting procedure is proposed for pose invariant geometry image and normal map extraction. The proposed 3D landmark detector employ the idea of statistical models (*i.e.*,

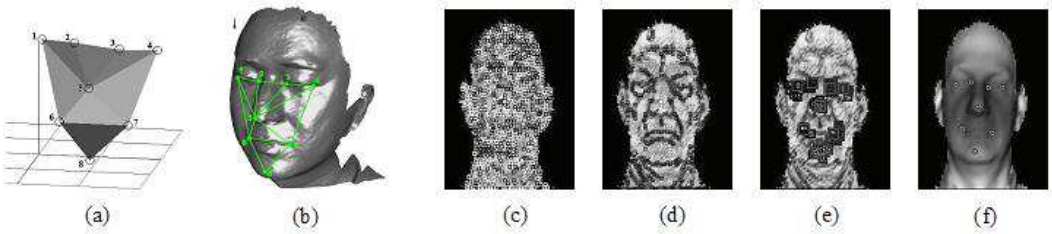


Figure 2.5: *Illustration of 3D landmark detection: (a) 8 points based Facial Landmark Model; (b) landmark model overlaid on a 3D facial surface; (c) extrema values of shape index; (d) spin image templet based filter; (e) landmark model fitting for consistent landmark selection; (f) selected best landmarks. Figures are taken from [Perakis et al. 2009] and [Passalis et al. 2011].*

Active Appearance Models (AAM) and Active Shape Models (ASM)) to build a generic Facial Landmark Model (FLM) using 8 anatomical landmarks of facial surface (Fig. 2.5 (a) and (b)). The mean shape of FLM is established by using *Procrustes Analysis* and the landmark shape variations is carried out by performing *Principal Component Analysis (PCA)*. All the landmark candidates are first filtered by extrema values of shape index (Fig. 2.5 (c)), and then filtered by the spin image templet (Fig. 2.5 (d)). Finally, the optimal combination of landmark set which has the minimal fitting error to FLM (Fig. 2.5 (e)) is selected as the best landmark set (Fig. 2.5 (f)). Once the landmark detection established, the pose of face data (*e.g.*, frontal, left side or right side) can be estimated. Meanwhile, the initial landmark based coarse registration and the final *Simulated Annealing* based fine registration between face data and the AFM can be done. The following symmetric deformable model fitting procedure is established by elastically deforming the left, right or both sides of AFM to the ones (left, right, or both sides) of left scan, right scan, or frontal scan, respectively. Then, the missing sides are recovered by symmetrically mirroring. These fitted AFM are used

to generate the pose invariant geometry and normal images. Finally, wavelet analysis is performed on those images for facial feature extraction and face matching.

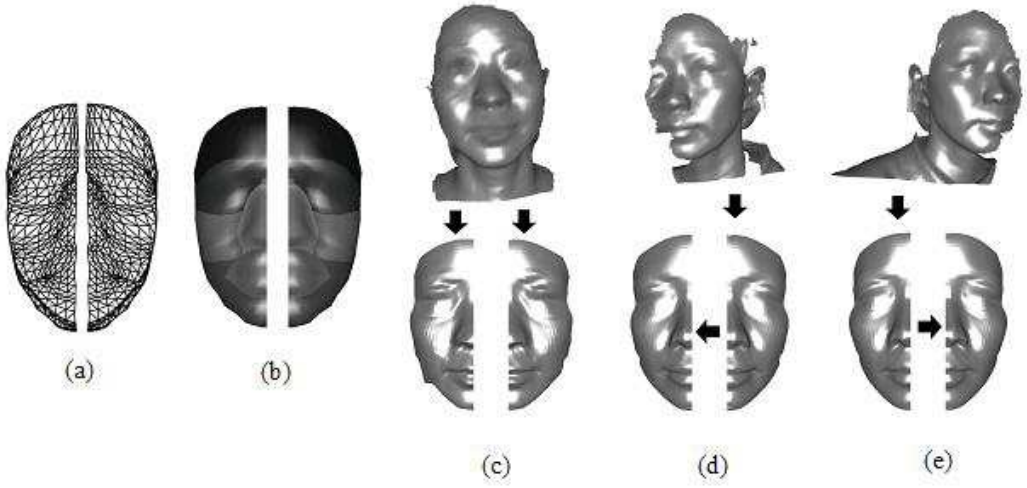


Figure 2.6: *Illustration of symmetric deformable model fitting: (a) left and right sides of the control mesh of AFM; (b) left and right sides of the annotated AFM; (c)-(e) raw data (top row) and fitted AFM (bottom row) of a subject with different poses. (c) both sides of the frontal scan are fitted; (d) left side of the left scan is fitted (and the right is mirrored); (e) right side of the right scan is fitted (and the left is mirrored). Figures are taken from [Perakis et al. 2009] and [Passalis et al. 2011].*

Based on the fact that the human face is not perfectly symmetrical, the main assumption for the authors of using facial symmetry is: the shape difference caused by facial asymmetry between the left and the right region of a person's face is less than the shape difference between these regions and the regions of another person's face. Their ideas are evaluated on a combined databases, collecting frontal scans from the FRGCv2.0 and side scans (yaw 45°, 60° and 90°) from the 3D Ear database of the University of Notre Dame (UND). The rank-one scores of 86.4%, 81.6% and 76.8% are reported on UND45, UND60 and UND90 databases, respectively. Meanwhile, a time cost of 18s for the whole pipeline is reported on a typical PC. As pointed out by the author, one limitation of their method is that at least half part of the face should be existing ($\leq 90°$ yaw rotation), so that facial symmetry can be used. On the other side, their current performances are largely blocked by the accuracy of the automatic facial landmark detector.

In [Chu *et al.* 2013], a new scenario (the gallery scan comes from one side of a face while the probe scan comes form the other) of partial face recognition is proposed. This new scenario is more challenging compared to the one of frontal gallery scan and partial probe scan. An improved version of the *Semi-Coupled Dictionary Learning* technique is used to modeling this problem and promising performances are achieved.

2.1.3 PCA Subspace based Deformation Modeling

Mian *et al.* [Al-Osaimi *et al.* 2009] propose a non-rigid 3D face face recognition approach by modeling an Expression Deformation Model (EDM) based on the *Principle Compo-*

nent Analysis (PCA) technique. The key idea of this method is that: among all the shape residues between a probe scan and all gallery scans, there is only one representing the pure expression deformations, the one in which the probe and gallery belongs to the same subject. The probe can be correctly recognized because the pure expression deformations can be reconstructed with minimal error if the expression deformation patterns can be well modeled. The model training and face matching stages are summarized as follows.

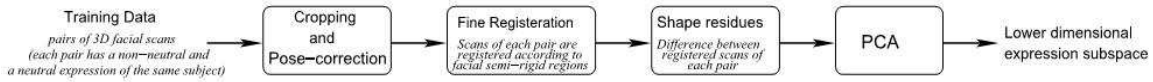


Figure 2.7: *Pipeline of training EDM. [Al-Osaimi et al. 2009]*

In the training stage, EDM is learned in a PCA subspace, which is built form shape residues between neutral and non-neutral scan pairs of the same subject. As shown in Fig. 2.7, the training data contains multiple non-neutral scans and one neutral scan for each subject. Ideally, to get a generic model, a wide range of facial expressions and a large number of subjects should be collected in the training data set. Moreover, to accurately compute shape residues, the scan pairs should be well-cropped, pose normalized, and finely registered. Formally speaking, let $\mathbf{S}_i$ and $\tilde{\mathbf{S}}_i$ denote the neutral and non-neutral range images of the i-training pair, Their shape residue is simply their differences,

$$\mathbf{R}_i = \tilde{\mathbf{S}}_i - \mathbf{S}_i.$$

Suppose there are totally n training pairs, and their covariance matrix is

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{r}_i - \bar{\mathbf{r}}_i)(\mathbf{r}_i - \bar{\mathbf{r}}_i)^T$$

where $\mathbf{r}_i \in \mathbb{R}^m$ is the vectorial expression of $\mathbf{R}_i$ and $\bar{\mathbf{r}}_i$ is the average shape residue. Then, the PCA subspace based expression deformation model $\mathbf{V} \in \mathbb{R}^{m\times k}(k \ll m)$ is defined by k eigenvectors of Σ

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k],$$

where $\Sigma\mathbf{v}_i = \lambda_i\mathbf{v}_i$, and λ_i is the k-th largest eigenvalue. Some eigenvectors of the matrix Σ are shown in Fig. 2.8.

Figure 2.8: *Some eigenvectors which span the shape residue PCA subspace (expression deformation model). the images from left to right are the 1st, 2nd, 3rd, 9th, 10th, 11th and 25th.eggenvectors, respectively. Notice that the eigenvectors with the higher eigenvalues are more relevant to the largely deformable regions (e.g. the mouth) and the other regions (e.g. eyebrows, eyes) gradually appear in the lower eigenvectors.[Al-Osaimi et al. 2009]*

In the matching stage, a probe $\mathbf{P}$ and a set of galley scans $\mathbf{G}_{1,\ldots,N}$ are first well-cropped, posed normalized, and finely registered in the same way as the training phase. Then, a

set of shape residues $\mathbf{R}_{1,\ldots,N}$ and its vectorial expressions $\mathbf{r}_{1,\ldots,N}$ are computed. For face identification, each $\mathbf{r}_i$ is projected to the EDM $\mathbf{V}$ with minimal L_2 reconstruction error. The projection coefficient $\mathbf{c}_i$ has the following closed-from expression

$$\mathbf{c}_i = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\mathbf{r}_i,$$

then, the shape residue can be reconstructed by

$$\tilde{\mathbf{r}}_i = \mathbf{V}\mathbf{c}_i = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\mathbf{r}_i.$$

The gallery scan with the minimal reconstructed error is regarded to be the probe's match. Considering the outliers of the shape residues and the artifacts at the border of the eigenvectors, the model $\mathbf{V}$ should be further modified to avoid the term $\mathbf{V}^T\mathbf{V}$ to be an identity matrix.

This method achieves a very high performance with a short processing time. A rank-one recognition of 98.14% on the whole FRGCv2.0 database under the *neutral vs. all* experimental setting. And a time cost of 4 seconds (facial signature extraction) plus 100ms (matching per gallery) is reported based on an Intel Core 2 processor. The drawbacks of this method may arise from: 1. the selection of a suitable number of training samples with different expressions; 2. the carefully face cropping and registration step which impact the artifacts at the borders of faces and the outliers in the shape residues.

2.1.4 3D Morphable Model based Deformation Modeling

Amberg *et al.* [Amberg *et al.* 2008] introduce an expression-invariant 3D face recognition method by building and fitting an identity/expression separated 3D Morphable Model (3DMM) [Blanz & Vetter 1999] to shape data. The identity model can be described as

$$f(\alpha_n) = \mu + \mathbf{M}_n\alpha_n$$

where μ denotes the mean shape, M_n is the matrix of eigenvectors vectors and α_n is the model parameter.

The expression model can be described as

$$f(\alpha_n, \alpha_e) = \mu + \mathbf{M}_n\alpha_n + \mathbf{M}_e\alpha_e = \mu + \mathbf{M}\alpha$$

where $\mathbf{M} = [\mathbf{M}_n \quad | \quad \mathbf{M}_e]$, $\alpha = \begin{bmatrix}\alpha_n \\ \alpha_e,\end{bmatrix}$ and $\mathbf{M}_e$ denotes the expression matrix and α_e denotes the expression coefficients.

As shown in the columns 1-3 of Fig. 2.9, given a target face scan with neutral or non-neutral expression, it is iteratively fitted to the 3D Morphable Model by minimize the following cost function

$$f(\mathbf{R}, \mathbf{t}, \alpha) = \sum_i \|\mu_i + \mathbf{M}_i\alpha + \mathbf{t} - \mathbf{R}u_i\|^2 + \lambda\|\alpha\|^2$$

where μ_i and $\mathbf{M}_i$ denote the rows of the model which correspond to the i-th vertex by subscript i, and u_i denotes a candidate scan-to-model correspondence point. Based on the fitting results, one can normalize/remove the facial expressions, the expression neutralized

scan are used for face recognition. The similarity measurement between faces is computed by the angle between face parameters in Mahalanobis space. The experiments shows that, comparing to the ICP based face recognition method, the performance of the proposed method does not decrease for the neutral database (*e.g.* UND database) and can be improved for the expression database (*e.g.* GavabDB).

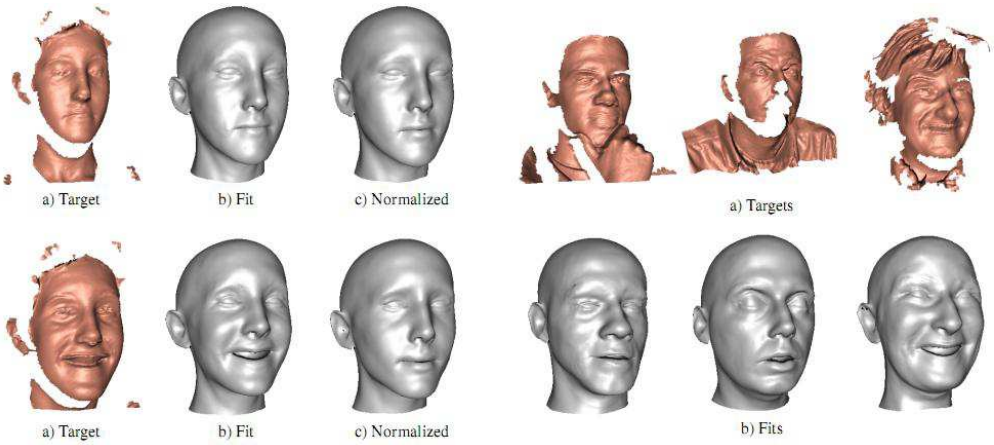


Figure 2.9: *Columns 1-3: Expression normalization of two scans of the same subject. The fitting algorithm gives a good estimate (b) of the true face surface given the noisy measurement (a). The pose and expression normalized faces (c) are used for face recognition. Columns 4-6: The reconstruction (b) is robust against scans (a) with artifacts, noise, and holes [Amberg et al. 2008].*

The main advantage of this method is that it is robust against scans with artifacts, noise, and holes (columns 4-6 of Fig. 2.9). For the disadvantages, this method does not report competitive recognition results on commonly used large scale 3D face databases (*e.g.* FRGCv2.0). Moreover, to build a 3DMM, a large number of facial landmarks are usually required to be manually annotated for getting dense correspondence as in [ter Haara & Veltkampb 2010].

2.1.5 Bilinear Model based Deformation Modeling

Similar to the spirit of 3D Morphable Model, Mpiperis *et al.* [Mpiperis *et al.* 2008a] propose to use a Bilinear Model [Tenenbaum & Freeman 2000] to encode identity and expression in independent control parameters to perform joint expression-invariant face recognition and identity-invariant facial expression recognition. The main procedures are depicted as in Fig. 2.10. Specifically, given two original 3D face scans (*i.e.* a gallery and a probe) of the same or different subjects displaying random expressions (first column), each face scan is approximately fitted by an general *landmark-guided* or *subspace-guided* elastically deformable model (second column). The general deformable model is created by performing subdivision on a base-mesh. The fitting produce is established by minimizing the total energy including *landmark distance*, *model-to-scan distance*, *scan-to-model distance*, and the elastic energy of the base-mesh. The aim of the deformable model fitting is to find the one-to-one point correspondence between *scan-to-model* and *scan-to-scan*. The fitted two scans have the same number of vertices and displaying the same poses. Thus, this step

can be regarded as pose and vertex normalization. Each normalized scan is then fitted to a symmetric bilinear model by minimizing the reconstruction square error, the reconstructed face scans are described in the third column of Fig. 2.10.

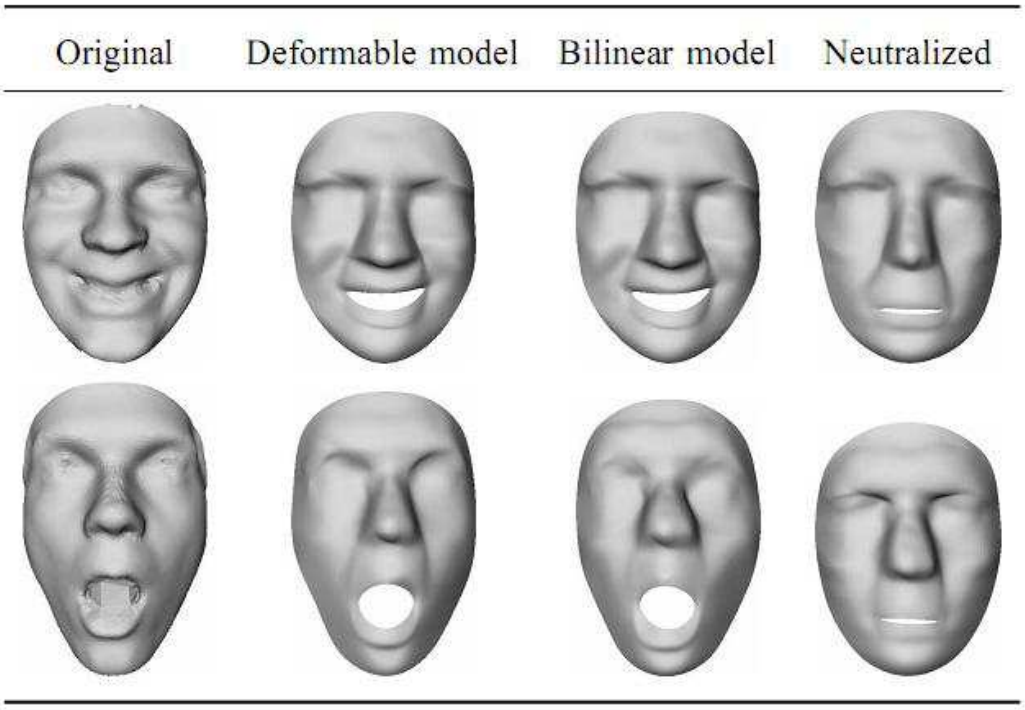


Figure 2.10: *Main procedures of the Bilinear Model based 3D face recognition [Mpiperis et al. 2008a].*

The bilinear model is built by a set of face scans in the training stage. Let $\mathbf{v}^{xp} \in \mathbb{R}^K$ be the vertices of facial surface of person p with expression x. Then each component $\mathbf{v}_k^{xp}$ is given by the general bilinear form [Tenenbaum & Freeman 2000]

$$\mathbf{v}_k^{xp} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ijk} a_i^x b_j^p$$

where a_i^x and b_j^p are the control parameters which control expression and identity respectively, while w_{ijk} are the coefficients which model the interaction of the factors.

Assume that there exist T faces belonging to T_p persons and each one displaying one of T_x possible expressions. The aim of training the bilinear model is to find the interaction coefficients w_{ijk} and the control parameters a_i^x and b_j^p for each identity-expression couple. Using the matrix notation, the bilinear model can be simplified to

$$\mathbf{v}_k^{xp} = \mathbf{a}^{xT} \mathbf{W}_k \mathbf{b}^p$$

where $\mathbf{a}^x = [a_1^x \cdot a_I^x]^T$, $\mathbf{b}^p = [b_1^p \cdot b_J^p]^T$ and $\mathbf{W}_k(i,j) = w_{ijk}$.

Once the bilinear model fitting procedure is done. The reconstructed scans are further neutralized by substituting the expression control vector $\mathbf{a}^x$ by the corresponding vector of

neutral faces (column 4 of Fig. 2.10). Note that gallery faces are neutral faces in most cases. The final expression free comparison between faces is established by the inverse of their vertex vectors' square Euclidean distance.

The performance of this method is evaluated on the BU-3DFE database, where half part (50 subjects) of the scans are used for training the elastically deformable model and computing the linear coefficients. The rest of the data, is split into the gallery set that includes a single face scan per subject (neutral or non-neutral), and probe set that includes the scans to be classified (various facial expressions). An average rank-one face recognition rate of 86% is reported, which shown to be better than Bronstein *et al.*'s [Bronstein *et al.* 2007] canonical images and a PCA-based method. Although promising results are achieved, as pointed out by the authors, there are still some issues that limit the performance: 1. The main limitation is the requirement of a large (universal) bootstrap set in which the facial expression should be correctly annotated; 2. The accuracy of the elastically deformable model fitting procedure is another key factor affecting the performance.

2.2 Surface-distance based Approaches

Surface-distance based approaches generally assume that facial surfaces are two-dimensional, compact, connected, zero-genus manifolds embeded in $\mathbb{R}^3$. Based on this assumption, surface geodesics, induced by the Riemannian metric, as an intrinsic geometry quantity, are used to construct expression-invariant representations of faces. This idea has been deeply discussed by several research groups (*e.g.* Bronstein *et al.*, Mpiperis *et al.*, and Berretti *et al.*). Furthermore, if facial surfaces are also differentiable, their tangent spaces are also manifolds. Thus, elastically matching of facial curves can be accomplished by inducing specific elastic Riemannian metrics on various facial curve-based shape spaces. These ideas have been successfully proposed and developed by Daoudi's group. According to different surface parts used for face matching, surface-distance based approaches are simply classified into three categories in our thesis: (i) iso-metric surface based approaches, which means the whole isometric approximated surfaces are used for face matching. (ii) iso-level curves based approaches, which means iso-depth curves, iso-geodesic curves or iso-geodesic path curves are used for facial features and for the final face matching. (iii) iso-geodesic strips based approaches, which means face matching is performed based on a set of equal width iso-geodesic strips. Their detailed descriptions are introduced as follows.

2.2.1 Iso-metric Surface based Approaches

In [Bronstein *et al.* 2003], Bronstein *et al.* explore an expression-invariant representations of faces for expression-invariant 3D face recognition. The expression-invariant representations of faces are obtained under the assumption of the isometric model. The isometric model assumes that all the facial surfaces with various expressions of the same subject are approximately intrinsically equivalent (i.e., have the same metric structure). The key idea is: all the facial expression variations only change the extrinsic geometry of facial surface, while the intrinsic geometry is invariant and thus can be contributed to the subject's identity. Formally speaking, suppose a facial surface can be represented as a smooth compact connected 2-D Riemannian manifold, denoted by S. And the geodesic distances on S induced by the Riemannian metric are denoted by $d_S(p_i, p_j)$ for $p_i, p_j \in S$. A mapping

$\phi : S \to S'$ is called an isometry if

$$d_S(p_i, p_j) = d_{S'}(\phi(p_i), \phi(p_j)) \tag{2.25}$$

for all $p_i, p_j \in S$. In practice, S is sampled as a finite set of points $\{p_1, p_2, \cdots, p_N\}$ and the geodesic distances $d_S(p_i, p_j)$ is approximately computed by using the fast marching method (FMM) [Kimmel & Sethian 1998] and stored in a $N \times N$ matrix Δ. The isometric representation of S is further isometrically embedded into a low-dimensional Euclidean space $\mathbb{R}^m$ for the following face matching procedure. Formally, the isometric embedding mapping $\varphi : S \to \mathbb{R}^m$ is approximated by the discrete mapping

$$\varphi_N : (\{p_1, p_2, \cdots, p_N\} \subset S, \Delta) \to (\{q_1, q_2, \cdots, q_N\} \subset \mathbb{R}^m, D) \tag{2.26}$$

such that $d_S(p_i, p_j) \approx d_{\mathbb{R}^m}(q_i, q_j)$ for all $i, j = 1, \ldots, N$. $\Delta = (\delta_{ij}) = (d_S(p_i, p_j))$ and $D = (d_{ij}) = (d_{\mathbb{R}^m}(q_i, q_j))$ are two metric matrices. The surface $(\{q_1, q_2, \cdots, q_N\} \subset \mathbb{R}^m, D)$ is called *canonical form* of S. Once canonical forms are computed, the problem of non-rigid matching between S is simplified to rigid matching between canonical forms. Figure 2.11 shows some examples of canonical forms of faces with strong facial expressions.

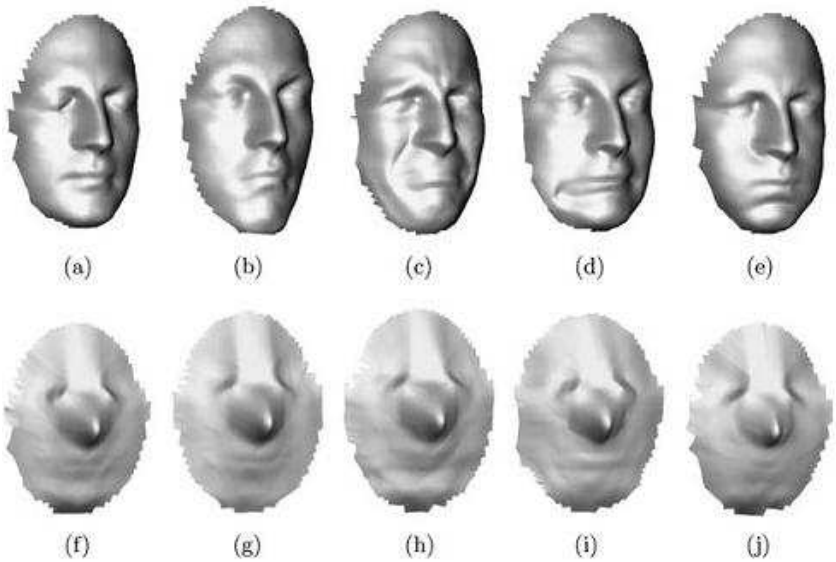


Figure 2.11: *Examples of canonical forms (f)-(j) of faces with strong facial expressions (a)-(e) [Bronstein et al. 2005].*

Different aspects of this isometric model has been extended by Bronstein *et al.* in their following works. In [Bronstein *et al.* 2004b], only metric tensor of the surface is used for geodesic distance computation, and a 3D face recognition method is performed directly on the normal fields of facial surfaces acquired by photometric stereo techniques. In [Bronstein *et al.* 2004a], texture-based canonical images are created and combined with 3D canonical forms for multi-modal 2D+3D face recognition. These two works are further extended to [Bronstein *et al.* 2005], in which the choice of the optimal embedding space is discussed and the recognition of identical twins is demonstrated. The influence of the embedding geometry and dimensionality choice are deeply studied in [Bronstein *et al.* 2007]. In [Bronstein *et al.* 2006], the isometric-invariant similarity between faces is measured by

embedding one facial surface to other. Thus, this method has the ability to handle partially missing data.

An obvious drawback of their methods is the open mouth problem. In this case, the genus zero surface assumption is invalid, and the isometric-invariant property of geodesic distance is also invalid. Moreover, only a limited experimental results carried out on some small databases are reported.

Mpiperis *et al.* [Mpiperis *et al.* 2007] propose a similar idea for 3D face recognition based on the geodesic polar representation. Their method outperforms the method of [Bronstein *et al.* 2007]. Comparing to [Bronstein *et al.* 2007], a fast geodesic computation method is proposed. Meanwhile, open mouth problem is improved by mapping lips to a continuous region on the geodesic plane so that the mouth hole vanishes. As pointed out by the authors, this method has the following limitations: (i) the assumption of the isometry of surface deformations will be invalid for extreme facial expressions; (ii) the geodesic polar parameterization depends on solving a differential equation. The accuracy of the solution depends on the sampling or the resolution of the surface. (iii) The performance is also influenced by the accuracy of the automatic detection of nose and sellion.

2.2.2 Iso-level Curves based Approaches

Based on the powerful surface shape analysis theory and Riemannian geometry theory, Daoudi's group develop a series of 3D face recognition methods by shape analysis of various facial curves. Their methods start by assuming that a facial surface S is a two-dimensional differentiable, connected, zero-genus manifold and mainly contain three procedures: (i) facial curves extraction for facial surface representation; (ii) Riemannian analysis of facial curve spaces for facial curve comparisons; (iii) facial surfaces measuring for face recognition.

As shown in Table 2.1 and Figure 2.12, three types of facial curves, namely, iso-depth curves [Samir *et al.* 2006], circular curves (iso-level curves) [Samir *et al.* 2009] [Ballihi *et al.* 2012], and radial curves (iso-geodesic path curves) [Drira *et al.*] [Ballihi *et al.* 2012] [Drira *et al.* 2013] have been proposed for facial surface representations. Iso-depth curves are closed planar curves sampled from pose normalized 2.5D depth images. An iso-depth curve can be extracted by slicing a depth image S using a plane P_λ parallel to the xy-plan of S. Circular curves are closed space curves sampled from pose normalized 3D facial surfaces. A circular curve can be extracted by slicing a facial surface S using a sphere $M_\lambda^{p_0}$ centered at the nose tip p_0 with a given radius λ. Radial curves are open space curves sampled from pose normalized 3D facial surfaces. A radial curve can be extracted by slicing a facial surface S using a plan $P_\theta^{p_0}$ passing the nose tip p_0 and perpendicular to the xy-plan of S.

In practice, facial surfaces are sliced by fixed orders with respect to λ and θ. Thus, all three types of facial curves can be extracted in order. Moreover, for each extracted facial curve, the procedures of spline fitting, arc-length parameterization and resampling are used for postprocessing. Finally, a facial surface can be represented as a set of ordered and uniformly sampled facial curves. Thus, one-to-one correspondence across facial surfaces can be achieved at both curve-level and point-level. Figure 2.12 shows an example of the above facial curve based surface representations. Table 2.1 summarizes the definitions, extraction methods, and properties of these three types of facial curves based facial surface represen-

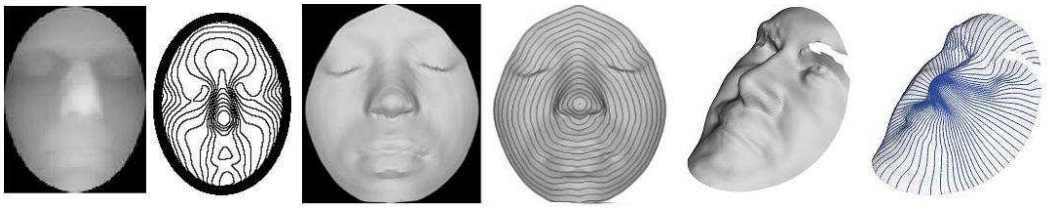


Figure 2.12: *Facial curves based surface representations: iso-depth curves [Samir et al. 2006], circular curves [Samir et al. 2009] and radial curves [Drira et al. 2013] (from left to right).*

Table 2.1: *Three types of facial curves based facial surface representations.*

Iso-depth curves	**Circular curves**	**Radial curves**
closed	closed	open
$\mathbb{R}^2$	$\mathbb{R}^3$	$\mathbb{R}^3$
$\alpha_\lambda = \{S \cap P_\lambda\}$	$\beta_\lambda = \{S \cap M_\lambda^{p_0}\}$	$\gamma_\theta = \{S \cap P_\theta^{p_0}\}$
$P_\lambda // S_{(x,y)}$	$M_\lambda^{p_0} = \{x \mid d(x,p_0) = \lambda\}$	$P_\theta^{p_0} \perp S_{(x,y)}$
-	$p_0 \in S$, nose tip	$p_0 \in S$, nose tip
$\lambda \in [\min z(S), \max z(S)]$	$\lambda \in [0, \infty)$	$\theta \in [0, 2\pi]$
$S \approx \cup_\lambda \alpha_\lambda$	$S \approx \cup_\lambda \beta_\lambda$	$S \approx \cup_\theta \gamma_\theta$
$\sim \mathbf{SO}(2)$	$\sim \mathbf{SO}(3)$	$\sim \mathbf{SO}(3)$

tations. It's easy to see that iso-depth curves based surface representation is invariant to planar rotation transformation group $\mathbf{SO}(2)$, and the other two types of representations are invariant to space rotation transformation group $\mathbf{SO}(3)$.

Once a type of facial curves (i.e. α_λ or β_λ or γ_θ) are extracted from a facial surface S, a pre-shape space $\mathscr{C}$ (i.e. a Riemannian manifold) is constructed with the elements of the velocity functions of curves (i.e. $\dot{\alpha}_\lambda$ or $\dot{\beta}_\lambda$ or $\dot{\gamma}_\theta$). Then, the quotient space $\mathscr{D}$ of the pre-shape space $\mathscr{C}$ is further defined rotation group $\mathbf{SO}(3)$ and the reparameterization group $\mathbf{S}^1$. The quotient space $\mathscr{D}$ is still a Riemannian manifold. Endowing it with a suitable (elastic) Riemannian metric, one can compute the geodesic path between two points in $\mathscr{D}$, which means the deformation process from one facial curve to another facial curve. The geodesic path can be computed by the shooting method for iso-depth facial curves and the path-straightening method for the other two types of curves. The length of the geodesic path supply an elastic measurement between two facial curves, which can be used as the similarity between two facial curves. Finally, the distance between two facial surfaces $d(S_1, S_2)$ is given by the average distances of all pairs of the corresponding facial curves laying on them. Please refer to Table 2.2 for a brief summary or refer to [Samir *et al.* 2006] [Samir *et al.* 2009] [Ballihi *et al.* 2012] [Drira *et al.*] [Ballihi *et al.* 2012] [Drira *et al.* 2013] for the detailed descriptions.

The main advantages of the iso-level curves based approaches are: (i) They can supply intrinsic surface based local coordinate systems for given facial surfaces; (ii) They can supply both curve-level and point-level one-to-one correspondence across facial surfaces; (iii) They can supply natural elastic metric between deformable and non-rigid facial surfaces.

Table 2.2: *Riemannian analysis of curve spaces for facial curves and surfaces comparisons.*

Facial curves and surfaces comparisons	Explanations
$\alpha_\lambda(s) : [0, 2\pi] \to \mathbb{R}^2$	iso-depth curves
$\dot{\alpha}(s) = e^{j\theta(s)}, j = \sqrt{-1}$	velocity function and angle function $\theta(s)$
$\mathscr{C} = \{\theta \mid \frac{1}{2\pi}\int_0^{2\pi}\theta(s)ds = \pi, \int_0^{2\pi} e^{j\theta(s)}ds = 0\}$	pre-shape space
$\mathscr{D} = \mathscr{C}/(\mathbf{SO}(2) \times \mathbf{S}^1)$	quotient space, rotation and reparameterization groups
shooting method	shooting by tangent direction g
$\langle g_1, g_2\rangle = \int_0^{2\pi} g_1(s)^T g_2(s)ds, g_1, g_2 \in T_q(\mathscr{D})$	Riemannian metric and tangent space
$\alpha_\lambda^1 \in S_1, \alpha_\lambda^2 \in S_2$	two iso-depth curves with same λ, from two surfaces
$\theta_\lambda^1 \in \mathscr{D}, \theta_\lambda^2 \in \mathscr{D}$	two points (angle functions) in $\mathscr{D}$
$d(\alpha_\lambda^1, \alpha_\lambda^2) = d_{\mathscr{D}}(\theta_\lambda^1, \theta_\lambda^2)$	length of the shortest geodesic path connecting θ_λ^1 and θ_λ^2
$d(S_1, S_2) = \frac{1}{n}\sum_{\lambda\in\Lambda} d(\alpha_\lambda^1, \alpha_\lambda^2), \lvert\Lambda\rvert = n$	distance between two facial surfaces
$\beta_\lambda(s) : [0, 2\pi] \to \mathbb{R}^3$	circular curves
$q(s) = \frac{\dot{\beta}(s)}{\sqrt{\lVert\dot{\beta}(s)\rVert}}$	square-root velocity function (SRVT)
$\lVert q(s)\rVert$	instantaneous speed
$\frac{q(s)}{\lVert q(s)\rVert}$	instantaneous direction
$\mathscr{C} = \{q \mid \int_0^{2\pi}\lVert q(s)\rVert q(s)ds = 0, \lVert q(s)\rVert = 1\}$	pre-shape space
$\mathscr{D} = \mathscr{C}/(\mathbf{SO}(3) \times \mathbf{S}^1)$	quotient space, rotation and reparameterization groups
path-straightening method	minimization of the length energy of the geodesic path
$\langle g_1, g_2\rangle = \int_0^{2\pi} g_1(s)^T g_2(s)ds, g_1, g_2 \in T_q(\mathscr{D})$	Riemannian metric and tangent space
$\beta_\lambda^1 \in S_1, \beta_\lambda^2 \in S_2$	two circular curves with same λ, from two surfaces
$q_\lambda^1 \in \mathscr{D}, q_\lambda^2 \in \mathscr{D}$	two points (velocity functions) in $\mathscr{D}$
$d(\beta_\lambda^1, \beta_\lambda^2) = d_{\mathscr{D}}(q_\lambda^1, q_\lambda^2)$	length of the shortest geodesic path connecting q_λ^1 and q_λ^2
$d(S_1, S_2) = \frac{1}{n}\sum_{\lambda\in\Lambda} d(\beta_\lambda^1, \beta_\lambda^2), \lvert\Lambda\rvert = n$	distance between two facial surfaces
$\gamma_\theta(s) : [0, 2\pi] \to \mathbb{R}^3$	radial curves
$q(s) = \frac{\dot{\gamma}(s)}{\sqrt{\lVert\dot{\gamma}(s)\rVert}}$	square-root velocity function (SRVT)
$\mathscr{C} = \{q \mid \lVert q(s)\rVert = 1\} \in \mathbb{S}^2$	pre-shape space
$\mathscr{D} = \mathscr{C}/(\mathbf{SO}(3) \times \mathbf{S}^1)$	quotient space, rotation and reparameterization groups
path-straightening method	minimization of the length energy of the geodesic path
$\gamma_\theta^1 \in S_1, \gamma_\theta^2 \in S_2$	two radial curves with same θ, from two surfaces
$q_\theta^1 \in \mathscr{D}, q_\theta^2 \in \mathscr{D}$	two points (velocity functions) in $\mathscr{D}$
$\psi(s) = \frac{1}{\sin\xi}(\sin((1-s)\xi)q_\theta^1 + \sin(\xi s)q_\theta^2)$	great circle, shortest geodesic path connecting q_θ^1 and q_θ^2
$\xi = d_{\mathscr{D}}(q_\theta^1, q_\theta^2) = \cos^{-1}(\langle q_\theta^1, q_\theta^2\rangle)$	length of geodesic path
$d(\gamma_\theta^1, \gamma_\theta^2) = d_{\mathscr{D}}(q_\theta^1, q_\theta^2)$	length of the shortest geodesic path connecting q_θ^1 and q_θ^2
$d(S_1, S_2) = \frac{1}{n}\sum_{\theta\in\Lambda} d(\gamma_\theta^1, \gamma_\theta^2), \lvert\Lambda\rvert = n$	distance between two facial surfaces

A rank-one recognition of 97% is reported for the facial radial curves based method on the whole FRGCv2.0 database [Drira *et al.* 2013].

The major disadvantages of the iso-level curves based approaches are: (i) The accu-

racies of the facial curve extractions are highly depending on the accuracy of the nose tip detection. Moreover, it will be problematic when the nose tip or the nose region of a facial surface is missing. A possible solution is to register this facial surface to a gallery model, and then use the nose tip of the gallery model for curve extractions [Drira *et al.* 2013]. However, when a larger part of a facial surface is missing (e.g. the profile of a facial surface), the common registration method (e.g. ICP) will be failed; (ii) Both the geometric shapes and the topological structures of the facial curves will be largely changed for the open mouth problem although this impact can be weakened by filling the hole of the opened mouth for the case of radial curves based representation [Drira *et al.* 2013].

2.2.3 Iso-geodesic Strips based Approaches

The iso-geodesic strips based approaches [Berretti *et al.* 2006], [Berretti *et al.* 2008], [ter Haar *et al.* 2008], [Berretti *et al.* 2010a] are proposed and developed by Berretti's group. As shown in Figure 2.13, the main procedures of their approaches can be summarized as follows: (i) Facial surfaces are partitioned into equal width and equal number geodesic stripes and sub-strips; (ii) The inter- and intra- distances between each pairs of strips (or sub-strips) of the same facial surface are computed by using 3D Weighted Walkthroughs (3DWW); (iii) A facial surface is represented as a graph, in which nodes represent strips (or sub-strips), and edges represent distances between strips (or sub-strips); (iv) Two facial surfaces are matched by matching two graphs.

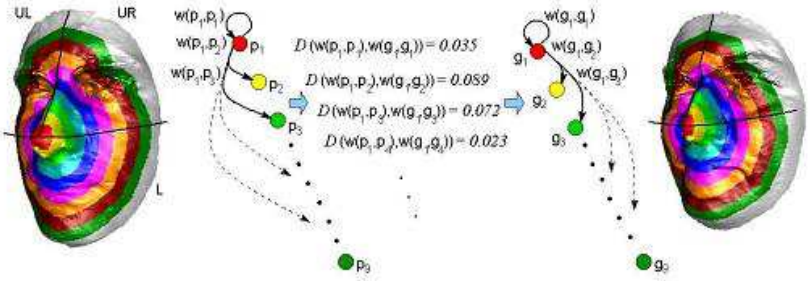


Figure 2.13: *Illustration of the main procedures for the iso-geodesic strips based approaches: (i) Each face is experimentally partitioned into 9 iso-geodesic stripes, and each strip is further partitioned into 3 parts: lower (L), upper-left (UL) and upper-right (UR) with respect to the coordinate of nose tip. In this figure, the UL parts are expressed by p_1 to p_9 for the left face and g_1 to g_9 for the right face. (ii) The inter-strip distances (i.e. $w(p_i,p_j), i \neq j$, $w(g_i,g_j), i \neq j$) and the intra-strip distances (i.e. $w(p_i,p_j), i = j$, $w(g_i,g_j), i = j$) are computed by using 3D Weighted Walkthroughs. (iii) Faces are compactly represented as connected graphs. (iv) Matching two faces by matching two graphs. The figure is taken from [Berretti et al. 2010a].*

The key concept of their approaches is the 3D Weighted Walkthroughs, which captures the mutual spatial displacement or relative position relationships (*left/right, up/down, front/back*) between two sets of points in $\mathbb{R}^3$. Formally, suppose $a = (x_a, y_a, z_a)$ and $b = (x_b, y_b, z_b)$ are two points in a 3D Cartesian coordinate system. Their mutual spatial

displacement can be encoded by the combinations of indexes $\langle i, j, k \rangle$ according to

$$i = \begin{cases} -1 & x_b < x_a \\ 0 & x_b = x_a \\ +1 & x_b > x_a \end{cases} \quad j = \begin{cases} -1 & y_b < y_a \\ 0 & y_b = y_a \\ +1 & y_b > y_a \end{cases} \quad k = \begin{cases} -1 & z_b < z_a \\ 0 & z_b = z_a \\ +1 & z_b > z_a \end{cases} . \tag{2.27}$$

The indexes $\langle i, j, k \rangle$ are defined as the 3D *walkthrough* from a to b.

Now given two iso-geodesic facial strips (or sub-strips) A and B, the number of unique pairs (a, b) with $a \in A$ and $b \in B$ that are connected by the same walkthrough $\langle i, j, k \rangle$ can be measured by the following (integral measure) expression:

$$w_{i,j,k}(A, B) = \frac{1}{K_{i,j,k}} \int_A \int_B C_i(x_b - x_a) C_j(y_b - y_a) C_k(z_b - z_a) dA dB, \tag{2.28}$$

where $dA = dx_a dy_a dz_a$ and $dB = dx_b dy_b dz_b$, $C_{\pm 1}(\cdot)$ are the characteristic functions of the positive and negative real semi-axes $(0, +\infty)$ and $(-\infty, 0)$, respectively, and $C_0(\cdot)$ is a quasi-everywhere null function. $\frac{1}{K_{i,j,k}}$ is a normalization factor and guarantees that $w_{i,j,k} \in [0, 1]$. Weights $w_{i,j,k}$ can be stored in a $3 \times 3 \times 3$ matrix $w(A, B)$, which is called 3DWW. In practice, the integral measure in Eqn. 2.28 can be efficiently computed by *directional indices*. The time complexity of computing the 3DWW between A and B is $O(n_A + n_B)$, where n_A and n_B being the number of points in A and B, respectively. Please refer to [Berretti *et al.* 2010a] for more details.

The authors claim that their approaches are robust to facial expression variations mainly for two reasons: (i) the property that iso-geodesic strips do not change too much under facial expressions; (ii) the continuity property of 3DWWs, which means slight point-drifting or deformations of two facial strips yield slight changes in their 3DWW. The authors report the best rank-one recognition rate (99.53%) on the SHREC08 face database, and 94.1% on the FRGC v2.0 database. Furthermore, the robustness to missing data, face cropping and the position of the fiducial points are also discussed by the authors. Similar to the isometric and iso-level curves based approaches, the open mouth problem is also a major drawback for their approaches.

2.3 Local Region/Feature based Approaches

Local region based approaches are generally proposed to handle local shape deformations caused by facial expression variations. This kind of approaches first divide the face into a number of regions (using some facial landmarks), then perform recognition (e.g. feature extraction, matching) on the separate regions, and finally fuse all the results. In this section, we will review several representative publications belong to this kind of approaches. They are Change *et al.* [Bowyer *et al.* 2005] [Chang *et al.* 2006], Faltemier *et al.* [Faltemier *et al.* 2008a], Alyuz *et al.* [Alyuz *et al.* 2010], Queirolo *et al.* [Queirolo *et al.* 2010], and Spreeuwers [Spreeuwers 2011].

In this thesis, local feature based approaches are further classified into two categories: (1) local geometry-texture descriptor based approaches; and (2) local geometry-shape descriptor based approaches. Local geometry-texture descriptor based approaches generally treat the range image as a 2D image in order to feature extraction. This kind of approach

has the advantage that a considerable number of well tested 2D texture descriptors for image processing can be directly applied. For example, Gabor filter, Log-Gabor filter, Haar wavelet, and Discrete Cosine Transform, see *e.g.* Cook *et al.* [Cook *et al.* 2006], Cook *et al.* [Cook *et al.* 2007], Wang *et al.* [Wang *et al.* 2010]; Local Binary Patterns (LBP) and eLBP, see *e.g.* Wang *et al.* [Wang *et al.* 2010], Huang *et al.* [Huang *et al.* 2011a] [Huang *et al.* 2012];

Local geometry-shape descriptor based approaches generally treat the range image as a free-form surface in order to feature extraction. This kind of approach has the advantage that a considerable number of well tested 3D shape descriptors for 3D object recognition can be directly applied. Moreover, most of the geometry-shape descriptors use the intrinsic geometric attributes of the surface, thereby they are pose-invariant. Some representative local geometry-shape descriptors used for 3D face recognition will be shortly reviewed in this section. They are including curvature-based descriptors, see *e.g.* Gordon [Gordon 1992], Tunwku and Ikeda [Tanaka & Ikeda 1996] [Tanaka *et al.* 1998], Li *et al.* [Li & Zhang 2007] [Li *et al.* 2009]; spin image [Conde *et al.* 2006]; point signature [Chua & Jarvis 1997]; local shape map [Wu *et al.* 2004]; 3D shape context [Berretti *et al.* 2011b]; spherical face representation [Mian *et al.* 2007]; tensor-based face representation [Mian *et al.* 2006].

2.3.1 Multiple Region Matching based Approaches

Change *et al.* [Bowyer *et al.* 2005] [Chang *et al.* 2006] propose a expression-robust expression recognition approach by matching multiple overlapping nose regions. These regions include a nose circle, a nose ellipse, and a region composed of just the nose itself (Fig. 2.14). Specifically, given a raw 3D face scan with registered 2D face image, a 2D skin detector is performed on the 2D image to remove the non-skin regions, and the corresponding face region is extracted. Then, curvature-based facial landmark detection method is adopted to detect the nose tip (peak region), eye cavities (pit region), and nose bridge (saddle region), where a quadratic surface fitting method is used for curvature estimation. The detected landmarks are used to extract gallery/probe local regions. As depicted in Fig. 2.14, three different regions are extracted around the nose tip for a probe, and a relatively large region is extracted around the nose tip for a gallery. Pose normalization of each local region is accomplished by aligning it to a generic 3D face model. This method uses the *ICP*-based matching algorithm to compare each gallery-to-probe region pair, and the *Mean Square Error (MSE)* measure is used to compute their similarity score. For classifier fusion, *product* and *sum* rules achieved the best performance. On the FRGCv2 *superset* database with a gallery of 449 subjects with a single neutral scan per person, the reported rank one recognition rates are 97.1% and 87.1%, respectively, for neutral probe and non-neutral probe sets. However, as pointed out by the authors, there is more to be done to solve the problem of handling expression variations, since there is about a 10% degradation when going from matching neutral expressions to matching varying expressions. Two important findings of this paper are: 1. Using less of the face (*e.g.* nose region) can result in more accurate recognition even for the scenarios of neutral-to-neutral matching. 2. Fusion the scores of multiple overlapping regions can largely improve the performance.

Based on the above two findings, Faltemier *et al.* [Faltemier *et al.* 2008a] propose a region ensemble method for 3D face recognition. As shown in Fig. 2.15, this method ex-

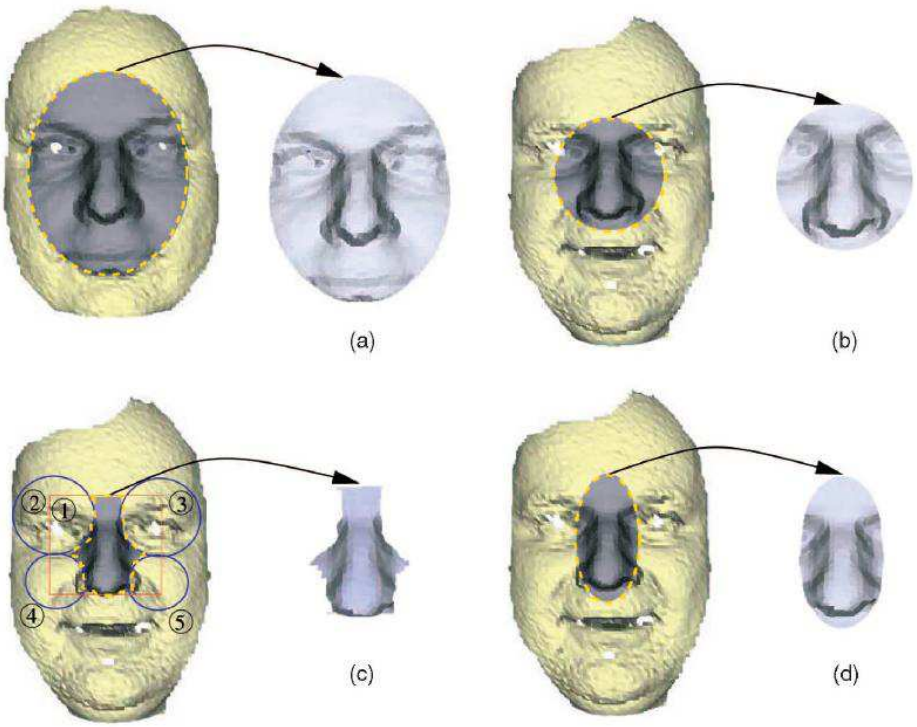


Figure 2.14: *Extraction of a gallery region and three probe regions. (a) A gallery region, (b) probe C (probe region in general center face area), (c) probe N (probe region in a nose region), (d) probe I (probe region in an interior nose region) [Chang et al. 2006].*

tract 38 regions distributed over the whole face. The centroids of these regions are defined using X and Y offset values from the nose tip location, and with a single radius or multiple different radii, resulting some overlapping regions. The nose tip is automatically detected by combing two or three methods discussed in their previous work [Faltemier *et al.* 2006]. The accuracy of the individual regions, the number of regions for fusion, and different types of fusion rules are discussed experimentally. Experimental results demonstrated that using 28 small regions on the face achieve the best performance, and the *Borda count* and *consensus voting* methods yield higher performance than the standard *sum, product*, and *min* fusion rules. They reported a rank-one recognition of 97.2% on the whole FRGCv2.0 database in a standard experimental protocol. As an obvious disadvantage, in all these studies, each probe region should compare to all the gallery faces based on the ICP-based matcher. This is a time-consuming task when the gallery set is large. For example, it takes about 1312 seconds for identification of a single probe using a gallery of 466 subjects of the FRGCv2 data.

Alyuz *et al.* [Alyuz *et al.* 2010] propose a regional registration based approach for expression-robust 3D face recognition. As shown in Fig. 2.16, this method consists of four parts: (1) Automatic landmark localization; (2) ICP-based global and local rigid face registration; (3) Facial feature extraction; (4) Region classifier fusion. In the first part, five fiducial points (nose tip, left/right inner eye pits and leftmost/rightmost points of the lower nose border region) are automatically located using curvature-based methods. These points are then used for coarse face alignment based on the *Procrustes Analysis*. The proposed region-

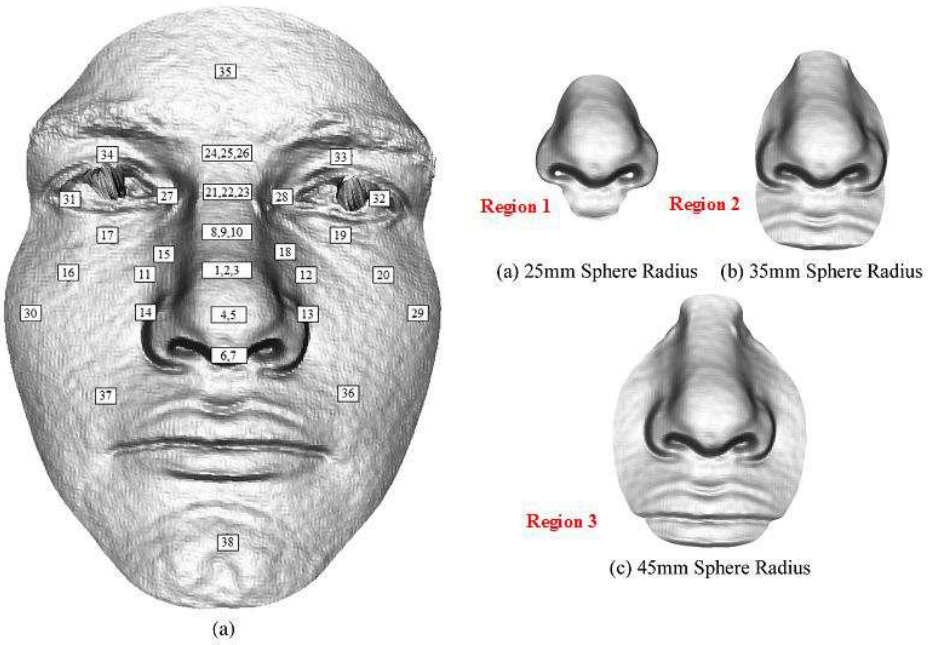


Figure 2.15: *left: Image of probe sphere centroids labeled by region number. Multiple region numbers at a centroid indicate that more than one radius was used for cropping, yielding multiple region probes with the same centroid. (a) 04514d324.* right: *Examples of nose regions with different radii [Faltemier et al. 2008a].*

based registration method is inspired by the average face model (AvFM)-based registration approach [Gokberk *et al.* 2006] [Gokberk *et al.* 2008] and extended to independent local regions, which is called the average region model (AvRM). The AvRM insists of 15 local patches, from which 7 meaningful regions (nose, left/right eye, forehead, left/right check, and mouth-chin) are constructed. Another region which covers local patches belonging to eye, nose, and forehead areas, in terms upperface, is also constructed. The AvRM-based registration provides one-to-one correspondence of all facial surface points. Two types of 3D shape descriptors: coordinates of facial surface points and curvatures of facial surface, are considered. The nearest neighbor algorithm is used to obtain the identification results for face classification. For region classifier fusion, both score-level (sum and product rules) and abstract-level (plurality voting) fusion approaches are considered. Two 3D face databases: the FRGCv2 and the subset of Bosphorus database excluding pose variation and occlusion variation scans are used for experimental evaluation. For experimental setup, the FRGCv2 database is divided into gallery and probe sets with 466 (first scan of each subject) and 3541 (remaining scans), respectively. The subset of Bosphorus is divide into galley and probe sets with 105 (neutral scan of each subject) and 2814 (remaining scans), respectively. For FRGCv2 database, the best rank-one recognition rate (94.80%) is achieved by fusing 16 individual classifiers (8 for each type of descriptor) with the modified plurality voting scheme. For Bosphorous database, the modified plurality voting scheme also achieves the best rank-one recognition rate (98.08%) by fusing two types of descriptors and 8 local regions. To further improve the performance, statistical analysis (LDA) is further performed on the coordinate based feature for dimensionality reduction. The achieved feature is called

statistical coordinate feature. The LDA subspace used for the FRGCv2 experiments is constructed using the FRGCv1 database, which consists of 943 3D face scans captured by the same sensor as FRGCv2. For Bosphorus experiments, 643 scans of 20 subjects of the subset of Bosphorus are used to construct LDA subspace and the 2265 scans of 85 subjects are

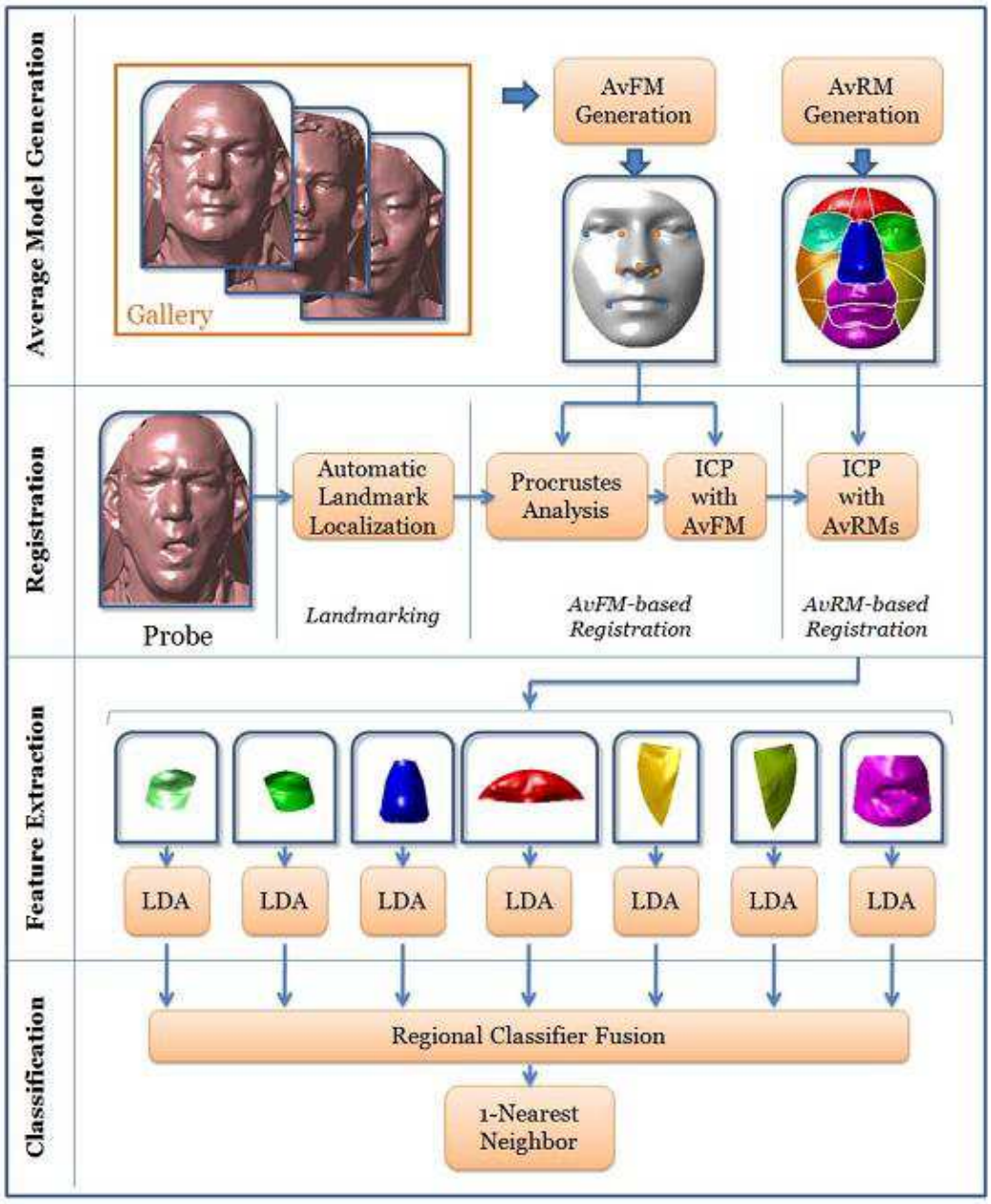


Figure 2.16: *Overview of the regional registration based 3D face recognition approach [Alyuz et al. 2010].*

used to form the gallery and probe sets. Base on these experimental setups, the rank-one recognition rates of 97.51% and 99.31% are reported on the whole FRGCv2 data and the

subset of Bosphorus data, respectively. Although high performance are achieved, the LDA approach has a poor generalization ability. The rank-one recognition rate on FRGCv2 will drop to 94.55% if the LDA space is constructed using the Bosphorus database. One important advantage of this method pointed out by the authors is its speed. Compare to the approaches of Change *et al.* [Bowyer *et al.* 2005] [Chang *et al.* 2006] and Faltemier *et al.* [Faltemier *et al.* 2008a], only a single registration is sufficient for a probe face benefiting from AvFM-AvRM approach. It totally takes about 131 seconds (11s for landmarking, 8s for curvature estimation, 10s for AvFM-based registration, 15s per region for AvRM-based registration, 3ms per region for similarity computation) for identification of a single probe using a gallery of 466 subjects of the FRGCv2 data.

Queirolo *et al.* [Queirolo *et al.* 2010] propose an automatic framework for 3D face recognition. Their method uses a *Simulated Annealing*-based approach (SA) for range image registration with the *Surface Interpenetration Measure* (SIM). As shown in Fig. 2.17, this method includes four main stages: *1. image acquisition; 2. preprocessing; 3. matching, and 4. evaluation.* Initially, the 3D face image can be captured by different 3D scanners. In the preprocessing stage, the face image is first smoothed with a median filter, and the face region is automatically detected by an ellipse detection method based on the Hough Transform. After the face region is detected, six landmark points, *i.e.* the inner left/right eye corners, the left/right nose corners, and the nose tip and base, are automatically detected for face segmentation. Three local regions: the circular area around the nose, the elliptical area around the nose, and the upper head, along with the entire face region are considered in the following face matching stage. More details about the preprocessing approach are described in their previous work [Segundo *et al.* 2007]. A Simulated Annealing-based approach (SA) is adopted to perform the matching process. SA-based matching process includes three main stages: *1) initial alignment, 2) coarse alignment, and 3) fine alignment.* The initial alignment is obtained by aligning two face images by their center of mass. The coarse alignment is performed using an SA-based searching process to minimize a robust alignment measure based on Maximum Likelihood Estimation SAmple Consensus (MLESAC) estimator [Torr & Zisserman 2000]. The fine alignment is established by performing an SA-based searching process to maximize a precise alignment measure based on Surface Interpenetration Measure (SIM) [Silva *et al.* 2005]. Finally, SIM scores of all the region pairs are fused by the *sum rule* to obtain the final similarity between two faces, which can be used for final face identification or verification.

The concept of Surface Interpenetration Measure (SIM) is first proposed in [Silva *et al.* 2005]. As illustrated in Fig. 2.18, in order to align two images, A and B, SIM is defined as

$$SIM_{(A,B)} = \frac{C_{(A,B)}}{|A|},$$

where the *interpenetrating point p in A with respect to B*, $C_{(A,B)}$ is defined as

$$C_{(A,B)} = \{p \in A | [(q_i - c) \cdot \vec{n}_c][(q_j - c) \cdot \vec{n}_c] < 0\},$$

where c is the nearest point of p in B, $\vec{n}_c$ is the local surface normal at c, $q_i, q_j \in N_p (i \neq j)$, and N_p denotes a neighborhood of p.

The authors performed extensive experiments on the FRGCv2 database to evaluate the performance of their method. In the identification scenario, a rank-one recognition rate

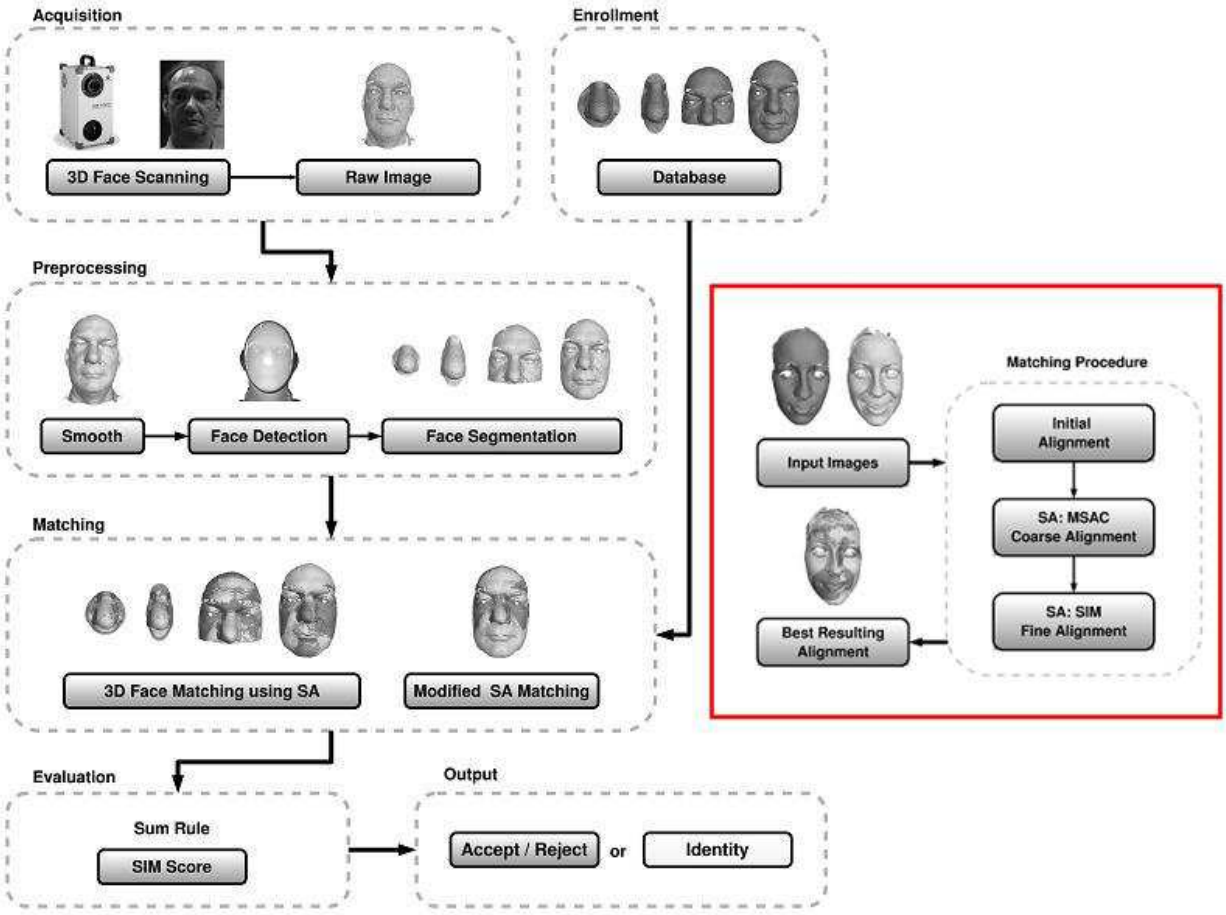


Figure 2.17: *The main stages of the Simulated Annealing based 3D face recognition framework and the main stages of the Simulated Annealing based face matching (in the red box) [Queirolo et al. 2010].*

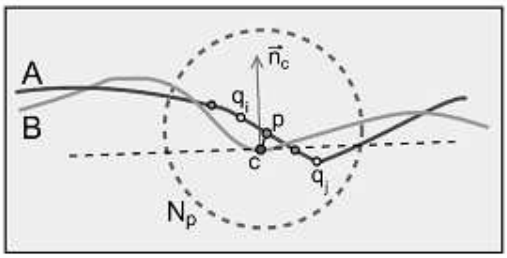


Figure 2.18: *Interpenetrating points p in A with respect to B [Silva et al. 2005].*

of 98.4% is reported. Compared to the ICP-based registration methods, this method has the following advantages: 1. It does not require prealignment of the range surface to a reasonable starting point. 2. It is robust to outliers arising from low surface overlap. Note that the proper convergence of ICP is guaranteed only if one surface is a subset of the other. 3. As a stochastic local searching algorithm, SA-based matching process has a

large probability to reach the global optimal solution. 4. Different from the Mean Squared Error (MSE) used in ICP-based algorithms, Surface Interpenetration Measure (SIM) is more precise to be used to assess the quality of the registration. One disadvantage of this method is that the SA-based registration procedure is quite time-consuming. It takes about 1864 seconds to identify a single probe using a gallery of 466 subjects of the FRGCv2 data [Spreeuwers 2011].

Different from those one-to-all local region registration based methods, e.g. Faltemier *et al.* [Faltemier *et al.* 2008a], Queirolo *et al.* [Queirolo *et al.* 2010], and the one-to-one regional face model registration based method, e.g. Alyuz *et al.*[Alyuz *et al.* 2010], Spreeuwers proposes a fast and accurate 3D face recognition method which performs face registration to an intrinsic coordinate system [Spreeuwers 2011]. The origin of the intrinsic coordinate system is defined as the nose tip, and the three angles of the coordinate system are determined by the symmetry plane and the angle of the nose bridge. The RANSAC (RANdom SAmple Consensus) algorithm [Fischler & Bolles 1981] is used to perform the cylinder fitting and the line fitting for face detection and the slope of nose bridge estimation, respectively. The aligned point cloud is used to generate high resolution range image by resampling, which is further normalized by performing spike removing, hole filling and face cutting. To handle expression issue, the authors defined a set of 30 overlapping regions which are shown in Fig. 2.19 where the white area is included and the black area is excluded. To compare two images, multiple region classifiers are fused to compute the

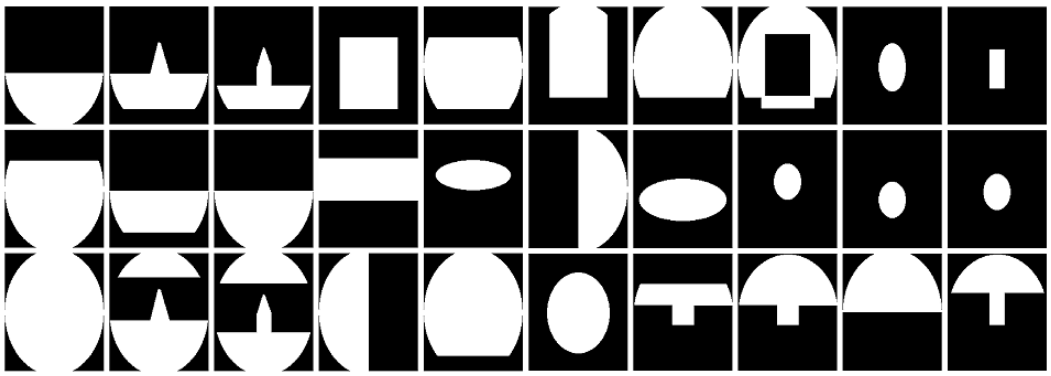


Figure 2.19: *30 local regions used for face comparison[Spreeuwers 2011].*

similarity. The vectorization form of each region is used to construct the feature vector, whose dimensionality is further reduced by performing PCA followed by LDA. Then, a likelihood ratio classifier is used for regional feature classification. Finally, the majority voting based fusion strategy is utilized for decision level fusion of all the region classifiers. For experimental evaluation, the Bosphorus database (all frontal scans without occlusions) and the 3DFace database are used for training of the region classifiers, while the FRGCv2 database is used for testing. Fig. 2.20 shows the rank-one scores for 30 region classifiers trained on the frontal Bosphorus database. The maximal rank-one score obtained for a single classifier is 95.9%, while a rank-one score of 97.9% is reported by fusing all the 30 classifiers. Similar results can be obtained for the region classifiers trained on the 3DFace database. Finally, the author reported a rank-one score of 99.0% by fusing two times 30

region classifiers trained on both databases.

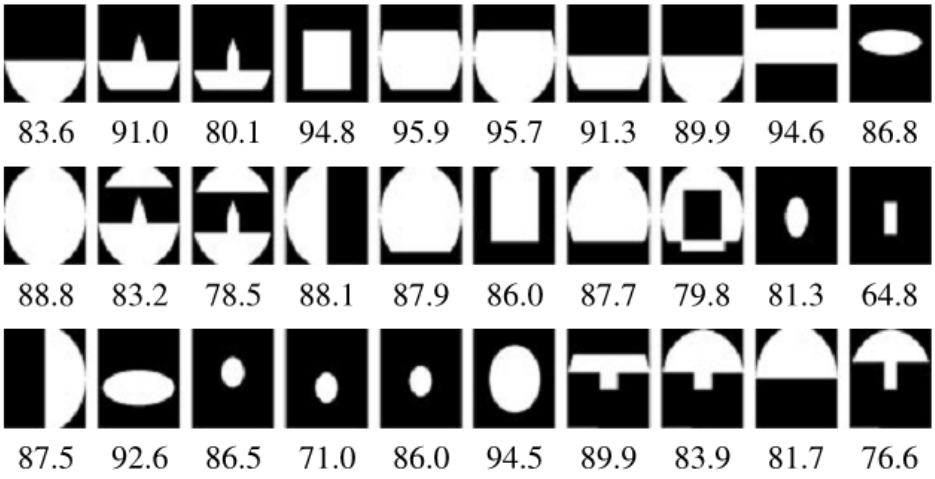


Figure 2.20: *Rank-one scores of individual region classifiers on the FRGCv2 database.*

Besides the high performance, the time cost is also a obvious superiority of this method. It takes only 2.5 seconds to identify a single probe using a gallery of 466 subjects of the FRGCv2 data, which is more than 700 times faster than the method in [Queirolo *et al.* 2010].

2.3.2 Local Geometry-Texture Descriptor based Approaches

Cook *et al.* [Cook *et al.* 2006] propose to perform Log-Gabor filter on range image to extract local facial descriptor for 3D face recognition. Compare to Gabor filter, Log-Gabor filter can capture more information in the high frequency areas and also has desirable high pass characteristics. Based on the fact that using multiple observations of a single face aids recognition performance, the authors use 18 Log-Gabor filters (6 orientations and 3 scales). Meanwhile, the face is broken into 49 square regions (7×7 grid) with 50% overlap in both horizontal and vertical directions. These regions are further decomposed by 3 scales of filter to generated 147 subregions. An overview of this process can be seen in Fig. 2.21. Principal Component Analysis (PCA) is then applied to the Log-Gabor filter responses in each of the 147 subregions for dimensionality reduction, and Mahalinobis Cosine distance measure is used for similarity computation. Finally, un-weighted sum rule score level fusion strategy is used for classifier combination and face recognition. Experimental results evaluated on the FRGCv2 database demonstrated the robustness of the proposed method. Rank-one scores of 92.93% is reported for the first *vs.* the remaining protocol and 94.63% for the first neutral *vs.* the remaining protocol.

This work is further extended to use multi-scale representation for 3D face recognition [Cook *et al.* 2007]. In [Cook *et al.* 2007], Haar wavelet transformation, Log-Gabor filter, and Discrete Cosine Transform (DCT) are used to extract multi-scale representations of 3D faces, respectively. Then, PCA and LDA techniques are used for dimensionality reduction. A standard nearest neighbor scheme with the mahalanobis cosine metric is employed, and the un-weighted summation of distances across all frequency bands is used as the final matching score. Two main conclusions are obtained: 1) A multi-scale partitioning of the

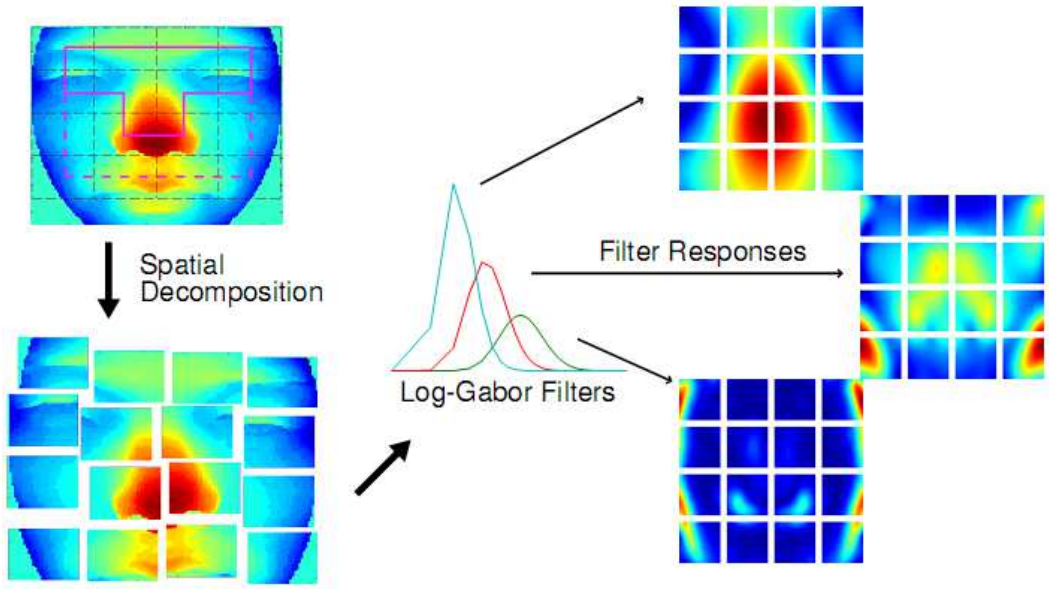


Figure 2.21: *Decomposition of the face into subregions.*

frequency domain can improve the discriminability of both the PCA and the LDA subspace projection techniques; and 2) the over-complete representation using Log-Gabor filters provides better partioning of the space/frequency domain for recognition then regular Gabor filters, discrete cosine transformation, or the wavelet transform.

Wang *et al.* [Wang *et al.* 2010] propose a 3D face recognition method, called *Collective Shape Difference Classifier* (CSDC). The framework of this method is shown in Fig. 2.22. This method consists of two phases: 1. *Offline Training*, and 2. *Online Testing*. Offline training phase includes three main stages: 1) *Preprocessing and Posture alignment*; 2) *Sign Shape Difference Maps (SSDMs) computation and description*; 3) *Boosting training of the Collective Shape Difference Classifier (CSDC)* ; While in the testing phase, the well-trained strong classifiers are employed for face verification and identification directly. In the first stage, input a raw face scan, three Gaussian filters are adopted to remove spikes, fill small holes, and smooth the data; a face region is detected on the texture channel, and the corresponding 3D points are labeled as face points; a plane is fitted using the face points, and the face is separated into two parts. The part with smaller mean square distance (MSD) from the points in the part to its centroid is selected as the candidate part for nose detection. In the candidate part, the farthest point to the fitted plane is selected as the approximate nose tip; facial region is cropped by a sphere centered at the nose tip. For posture alignment, facial symmetry plane is detected by registering a pair of mirrored face scans; Nose tip is precisely located on the facial central profile; Nose bridge direction is further estimated by fitting the points on the facial central profile; A standard coordinate system is defined by the nose tip, nose bridge direction, and the normal of the facial symmetry plane. All the facial points are transformed to this standard coordinate system for accurately posture alignment. The aligned face model is then projected to a normalized depth image. Given two depth images I_1 and I_2, their SSDM D_s is defined as $D_s(i,j) = I_1(i,j) - I_2(i,j)$. D_s captures not only shape differences, but also shape change patterns. The proposed Collective

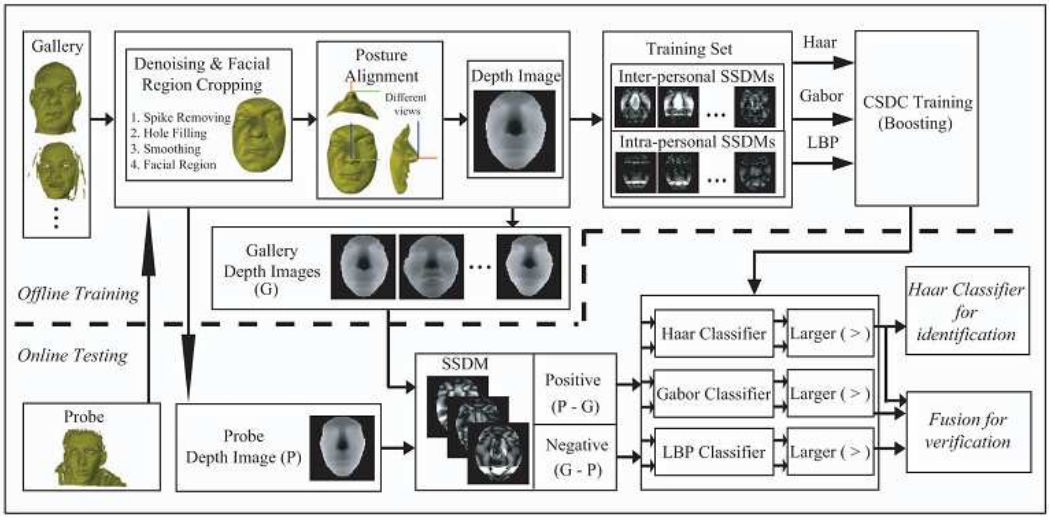


Figure 2.22: *The framework of Wang* et al.*'s method [Wang et al. 2010].*

Shape Difference Classifier (CSDC) is defined as $H_T(D_s) = \sum_{t=1}^{T} c_t(D_s)$, where D_s denotes a SSDM, $c_t(D_s)$ is a weak classifier selected based on the local features on the SSDMs during the boosting training. Three kinds of local features: Haar-like features, Gabor features, and multiblock Local Binary Patterns (LBPs) are used to describe different characteristics of the SSDM. Haar-like features measure the difference between region averages of the SSDM, which encode the change pattern of the shape difference; Gabor feature encode the characteristics of a set of spatial localities and orientations of the SSDM; LBP features capture the texture characteristics of the SSDM. To learn the CSDC, inter-personal and intra-personal SSDMs are built from given 3D face models. The BU-3DFE database are used as the training set. Finally, three CSDCs, H_T^{Haar}, H_T^{Gabor} and H_T^{LBP} are obtained. Comprehensive experimental results are reported on the FRGCv2 database. With the FAR of 0.1 percent, the authors report the verification rates of 97.97% for ROC I mask, 98% for ROC II mask, and 98% for ROC III mask. For face identification, only H_T^{Haar} is used. Rank-one scores of 98.3% and 98.4% are reported for the first *vs.* the remaining protocol and the first neutral *vs.* the remaining protocol, respectively. Besides the high recognition accuracy, this method also has a low computation cost. It takes about 3.6 seconds to recognize against a gallery with 1,000 faces. The main limitation of this method is that the procedure of pose alignment will fail when too many data points of the nose are missing, which is a common case for partial face recognition scenario.

Huang *et al.* extend their previous work [Huang *et al.* 2011e] [Huang *et al.* 2011a], and propose a 3D face recognition method using extended Local Binary Patterns (eLBP)-based facial description and local feature hybrid matching [Huang *et al.* 2012]. Input a raw face scan, this method first perform the basic preprocessing, e.g. spike removal, smoothing and hole filling. Then, eLBP descriptor is extracted. Different from the original LBP, eLBP not only computes the relative gray value difference between central pixel and its neighbors as provided by LBP, but also focuses on their absolute difference which are also important to describe the detailed information of local shapes. Similar to LBP, multi-scale extension of eLBP is used to generate multi-scale extended LBP. As shown in Fig. 2.23, 8 scales and 4

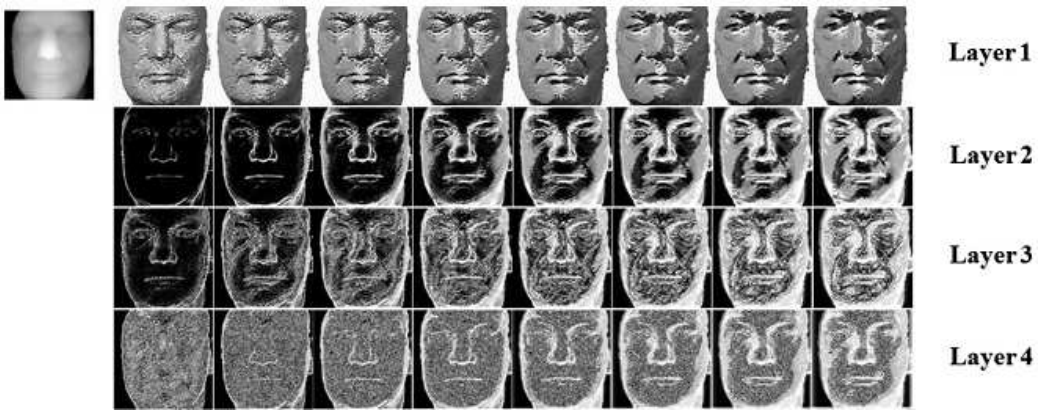


Figure 2.23: *MS-eLBP-DFs of a facial range image with different radii from 1 to 8 (from left to right), the layer 1 is the same as original LBP, from layer 2 to layer 4, the absolute difference between central pixel and its neighbors are encoded [Huang et al. 2012].*

layers are considered. Finally, each original depth image can be represented by 32 multi-scale eLBP depth faces (MS-eLBP-DFs). Once the MS-eLBP-DFs have been produced, the widely used SIFT features [Lowe 2004] are extracted from them for keypoints detection and description. Then, a hybrid matching process is performed, which combines a local matching step using SIFT-based features with a global one under the facial component and configuration constraints. Finally, three types of matching similarities: the number of matched keypoint pairs, similarity of the facial component constraint and similarity of the facial configuration constraint are fused using a basic weighted sum rule. This method is evaluated on three public databases: i.e. FRGCv2.0, Bosphorus, and Gavab DB. A rank-one score of 97.6% is reported on the FRGCv2.0 database using the neutral *vs.* all protocol, and a rank-one score of 97.0% is achieved on a subset of the Bosphorus database (3301 scans of 105 subjects with different expressions and occlusions). On the Gava DB database, the overall rank-one score of 91.39% is reported. All these experimental results demonstrate that the proposed method is robustness to facial expression variations, external occlusions, as well as extreme pose changes. One of the main advantage of this method is that the generally costly registration is not required thanks to the relative tolerance of the proposed hybrid matching algorithm to nearly frontal faces like the ones in the FRGCv2.0 and the subset of Bosphorus. However, to deal with extreme pose, e.g. left and right profiles, a coarse alignment step based on a few manually facial landmarks are needed.

2.3.3 Local Geometry-Shape Descriptor based Approaches

As early as 1990's, Gordon [Gordon 1992] introduced the outline of the use of curvature information in the process of 3D face recognition. Specifically, several aspects of using curvature information for face descriptions are considered: 1. *sign of Gaussian and mean curvature* is used to segment face into convex, concave, and two types of saddle regions (see Fig. 2.24); 2. *extrama values of two principle curvatures* are used to extract rigid and valley lines on face surface; 3. *principle directions* are used to detect umbilic facial points; 4. curvature maps are used to detect facial features such as nose region, nose bridge, eye

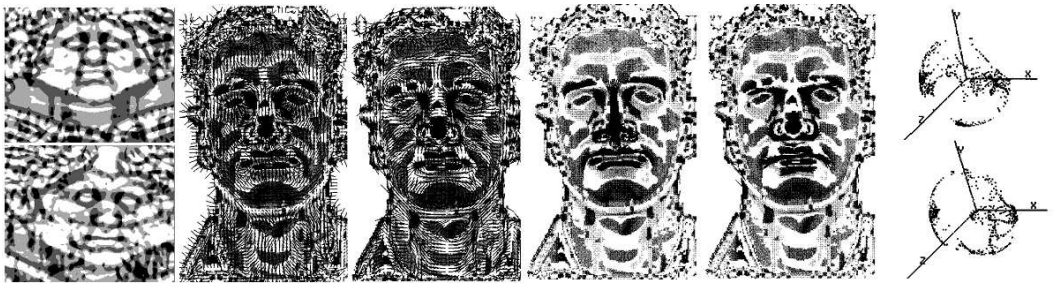


Figure 2.24: *From left to right: segmentation of two faces by sign of Gaussian and mean curvature: convex (white), concave (black), and two types of saddle: positive mean curvature (light grey), negative mean curvature (dark grey) [Gordon 1992]. maximum principle directions; minimum principle directions; rigid lines; valley lines; EGI of rigid lines (top), and EGI of valley lines (bottom) [Tanaka et al. 1998].*

corners, eyeballs etc. The detected facial features are then used in the process of face pose normalization. Finally, depth templates are compared to implement face recognition. Experiments are carried out on a test set of three views of each of 8 faces and recognition rates as high as 100% are reported. Similar idea is also explored by Tunwku and Ikeda [Tanaka & Ikeda 1996] [Tanaka *et al.* 1998], in which they propose to use *Extended Gaussian image (EGI)* of facial rigid and valley lines for face representation (See Fig. 2.24) and Fisher's Spherical Correlation as similarity measurement. Experimental results carried out on 37 range faces shows that shape information from surface curvatures provides vital cues in identifying human faces. However, these kinds of directly pointwise description and comparison of faces have obvious drawbacks: 1) In practice, the one-to-one point correspondence between two face surfaces is always unavailable; 2) The point-wise features usually have a very high dimensionality and contain a lot of redundant information. Fortunately, all these two drawbacks can be improved by the following two ways:

a) Establish one-to-one point correspondence between faces by fitting all of them to a generic face model, or by performing surface remeshing (or resampling). In [Alyuz *et al.* 2010], a generic face model is used. In [Li & Zhang 2007] [Li *et al.* 2009], Li *et al.* give an example of using surface remeshing technique. They explore the use of multiple intrinsic geometric attributes, including *angles, areas, geodesic distances, curvatures* for 3D face recognition. To obtain the one-to-one correspondence and compare those low-level features, each face is represented by a triangular mesh, 43 mask vertices are manually detected on it, and an uniform remeshing procedure is implemented (see Fig. 2.25). To deal with expression variations, they train different weights for each individual attribute [Li & Zhang 2007], or perform feature ranking [Li *et al.* 2009]. Based a Nearest Neighbor classifier [Li & Zhang 2007] and a Sparse Representation Classifier (SRC) [Li *et al.* 2009], they display the effectiveness of their methods on the Gavab database and the FRGC 3D database.

b) Construct compact local shape descriptors. Local shape descriptor, as a compact feature vector, designed to describe the local shape of a object by quantizing some kinds of basic geometric attributes (*e.g.* curvature, distance, area, angel) in a regular or parametric space, has been widely explored in the filed of 3D object recognition and 3D face recognition.

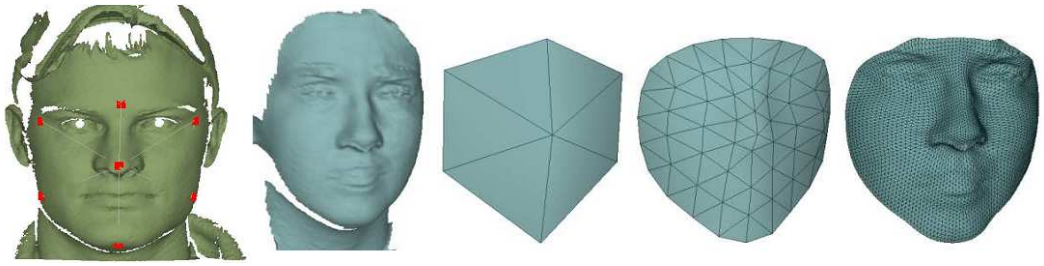


Figure 2.25: *Facial mask (left), An original face mesh and its Remeshings [Li et al. 2009].*

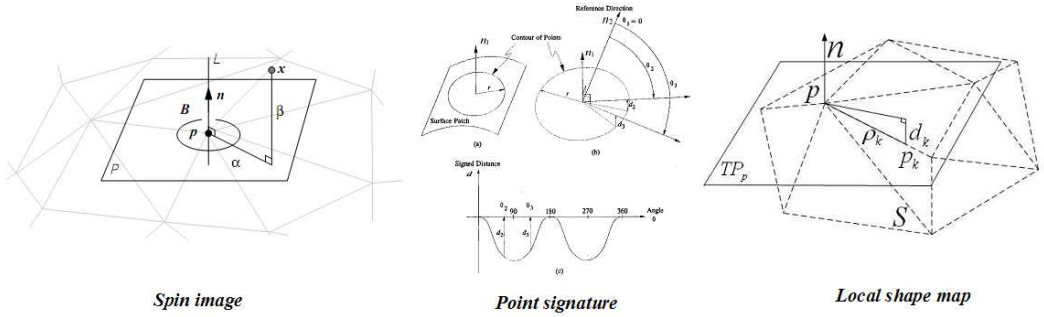


Figure 2.26: *Illustration of three local shape descriptors: spin image (left), point signature (middle), and local shape map (right).*

Spin image [Johnson & Hebert 1999], as a well-known 3D shape descriptor, has proven useful for face surface registration [Kakadiaris *et al.* 2007a], facial feature point extraction [Conde *et al.* 2006], and face recognition [Wu *et al.* 2004] [Mian *et al.* 2007]. Given a basic point and a support region of a surface mesh, a spin image of this point is generated by spinning an image (for example, of size 6×6) around the normal of the basic point and summing the surface points as they pass through the bins of the image (see Fig. 2.26).

Point Signature [Chua & Jarvis 1997] is proposed to describe the shape of the 3D contour curve of a basic point. Based on a predefined local coordinate system, a 2D planar contour curve is generated by projecting the 3D contour curve to the tangent plane of the basic point. Then, the point signature is defined by the sign distance from the two curves according to a constant angle sampling of the 2D planar curve (see Fig. 2.26). In [Chua *et al.* 2000], Chua *et al.* propose to use point signature for 3D face registration and recognition. Experimental results evaluated on 6 subjects, each with 4 different expressions, demonstrate the effectiveness of their method.

Local Shape Map [Wu *et al.* 2004] is motivated by the spin image and the point signature. It is a 2D histogram constructed by mapping 3D coordinates of surface's points with a sphere centralized at this point into 2D space. The distance from a basic point to its neighbor point and the displacement of the neighbor point to the tangent plane of the basic point are quantized to construct the 2D histogram (see Fig. 2.26). Testing on a database of 31 range images of 6 subjects, the authors claim that local shape map outperforms spin image for 3D face recognition.

3D Shape Context [Frome *et al.* 2004], as a straightforward extension of the original

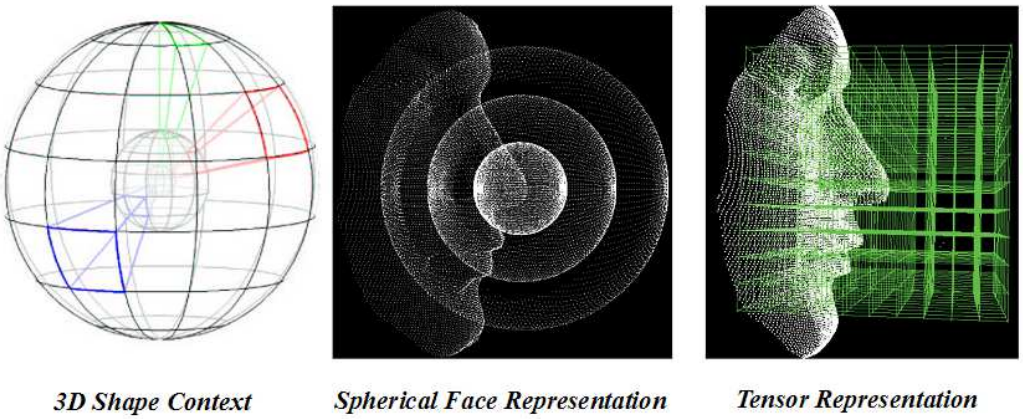


Figure 2.27: *Illustration of three local shape descriptors: 3D shape context (left), spherical face representation (middle), and tensor-based face representation (right).*

2D shape context [Belongie *et al.* 2002], describes the point distribution of a local shape by quantizing in a 3D spherical sampling space (see Fig. 2.27). In [Berretti *et al.* 2011b], Berretti *et al.* propose to use 3D shape context to describe the local shape of facial surface at the points detected by the SIFT descriptor. The effectiveness of their local shape descriptor is evaluated for the scenario of 3D partial face recognition.

Spherical Face Representation (SFR) is proposed by Mian *et al.* [Mian *et al.* 2007]. In which, an efficient multimodal 2D-3D hybrid face recognition approach is developed. SIFT descriptor is used to perform the 2D face matching, while for 3D face matching, both SFR descriptor and region-ICP (eye-forehead and nose) are used. A SFR can be considered as the quantization of the point cloud of a face into spherical bins centered at the nose tip. (See Fig. 2.27 as an example of 3 spherical bins). To compute an n bin SFR, the distance of all points from the nose tip is computed. These distances are then quantized into a 1-D histogram of $n+1$ bins. A rank-one recognition of 96.2% is reported by using the region-ICP algorithm. This score is further improved to 97.37% by fusing the 2D-SIFT matching, 3D-SFR matching and the region-ICP matching.

Tensor-based Face Representation is also proposed by Mian *et al.* [Mian *et al.* 2006]. It is defined as the quantization of the surface area of a face into a predefined 3D grid. Fig. 2.27 (right one) shows a $10 \times 10 \times 10$ tensor over a facial point cloud. In [Mian *et al.* 2008], Mian *et al.* propose another multimodal 2D-3D hybrid face recognition approach based on the framework of local keypoint detection, description, and matching. 2D face recognition is accomplished by performing 2D SIFT keypoint detection, description and matching procedures on texture images; while for range image based 3D face recognition, Principle Component Analysis (PCA) is used for local keypoint detection and local coordinate system assignment, then modified *tensor-based face representation* is employed for local shape description. The local shape features are further projected to a PCA subspace for dimensionality reduction. Finally, graph based matching strategy is explored for local feature matching. The proposed method achieves a rank-one recognition rate of 96.1% on the whole FRGCv2.0 database.

2.4 Discussion

Expression deformation modeling based approaches are originally proposed for dealing with the expression problem. Along with expression variations, matching of 2.5D probe scans in the presence large head pose changes ($\sim 45^{\circ}$ left from frontal and $\sim 45^{\circ}$ right from frontal) to full 3D gallery face models is also discussed in the TPS-based deformation modeling approach [Lu & Jain 2008]. But very limited experiments are displayed. As mentioned before, the annotated deformable model based deformation modeling framework has been extended to recognize 3D partial face caused by large head pose variations [Perakis *et al.* 2009] [Passalis *et al.* 2011] [Chu *et al.* 2013]. However, their methods are largely depending on the accuracy of 3D facial landmarking techniques. To the best of our knowledge, there are no published works to study 3D face recognition with missing data caused by large pose and external occlusion variations for other four kinds of expression deformation modeling based approaches.

For the category of surface-distance based approaches, it is visible that the iso-metric surface based approaches are originally proposed for constructing expression-invariant representations of faces [Bronstein *et al.* 2003]. To deal with partially missing data like simulated "big holes", iso-metric embedding among facial surfaces is also discussed [Bronstein *et al.* 2006]. But this idea does not test for real missing face data caused by head pose and external occlusion variations. On the other hand, approaches based on iso-depth curves [Samir *et al.* 2006] and circular curves [Samir *et al.* 2009] can deal with the face data only in the presence of a limited facial expressions of closed mouth; approaches based on radial curves can deal with the face data in the presence of not only various expressions, but also pose and occlusion variations [Drira *et al.* 2013]. To handle the missing data caused by pose variations, the authors introduce a restoration step by building a PCA-based statistical model on shape manifolds of curves in order to complete the partially-observed radical curves. To handle occlusions, the authors introduce an occlusion detection and removal step based on recursive-ICP. However, the nose tips in profile faces in the GavaD-B and occluded faces in the Bosphorus database should be annotated manually. For the iso-strip based approaches, 3D face recognition in the presence of only facial expression variations is evaluated.

Compared to the above two categories of holistic matching approaches, local feature based matching approaches have much more potential superiorities to handle expression, pose and occlusion variations under an uniform framework. For example, the multi-region matching approach proposed by Alyuz *et al.*[Alyuz *et al.* 2010] has been extended to handle the occlusions in their following works [Alyüz *et al.* 2012b] [Alyüz *et al.* 2012a] [Alyüz *et al.* 2013]. Their basic idea for handling occlusions is to improve the previous region-based registration strategy to a registration scheme which can adaptively select the non-occluded facial parts. The local geometry-texture descriptor based approach proposed by Huang *et al.* [Huang *et al.* 2011a] has also been extended to recognize the missing face data caused by pose variations [Huang *et al.* 2012]. As mentioned before, to deal with extreme pose, e.g. left and right profiles, a coarse alignment step based on a few manually facial landmarks is required for their approach.

The local geometry-shape descriptor based approaches can supply a promising solution to handle the puzzle problems of using manually annotated facial landmarks and the time-consuming registration algorithms, even for the cases of extreme pose and severe occlusion.

The main reason is that the geometry-shape descriptors are pose-invariant in nature. Specially, the framework proposed in [Mian *et al.* 2006], which combines local feature detection, description, and matching has been proved very powerful in the field of 3D face recognition [Maes *et al.* 2010] [Li *et al.* 2011c] [Berretti *et al.* 2011b] [Berretti *et al.* 2013] and 3D object recognition [Zaharescu *et al.* 2009] [Sun *et al.* 2009] [Darom & Keller 2012]. Recently, a large number of 3D keypoint detectors are proposed such as the Mesh-DoG 3D detector [Zaharescu *et al.* 2009], Heat Kernel Signature (HKS) 3D detector [Sun *et al.* 2009], Laplace-Beltrami operator 3D detector [Ruggeri *et al.* 2010], Harris 3D detector [Pratikakis *et al.* 2010] *etc.*, see [Tombari *et al.* 2013] for a comprehensive survey. Meanwhile, many 3D local shape descriptors are also developed such as Mesh-histogram of gradient (HOG) [Zaharescu *et al.* 2009], Scale-Invariant Spin Image [Darom & Keller 2012], Heat Kernel Signature (HKS) [Sun *et al.* 2009], mesh-SIFT [Maes *et al.* 2010], histogram of multiple-order surface differential quantities [Li *et al.* 2011c] *etc.* Due to that the 3D keypoint detectors can work automatically; the 3D local shape descriptors are pose-invariant in nature; and the local feature matching strategy has the ability to select and match the regions with less shape deformations, when used for 3D face recognition, this framework has a large potential to simultaneously deal with the challenges of facial expression variations [Maes *et al.* 2010] [Li *et al.* 2011c], data missing caused by head pose variations [Maes *et al.* 2010] [Li *et al.* 2011c] [Berretti *et al.* 2011b] [Berretti *et al.* 2013], and external occlusions [Maes *et al.* 2010] [Li *et al.* 2011c].

2.5 Summary

In this chapter, we have extensively reviewed three categories approaches: *i.e.* expression deformation modeling based, surface-distance based, and local region/feature based, for 3D face recognition. For each category, the most representative and successful approaches are deeply discussed. Among these approaches, the local region/feature based approaches are the most promising category to achieve high performance and to handle expressions, large pose and external occlusion variations in an uniform framework.

To achieve high performance, local region/feature based approaches usually describe face scans by multiple observations (*e.g.* multiple local regions, multiple encoding scales or multiple encoding parameters), different weights for different observations may be considered, then, the performance of all observations are well-fused to obtain the face similarity for the final decision. Based on this idea, we propose a local geometry-texture descriptor based approach in chapter 3.

To handle expressions, pose and occlusion in an uniform framework, as discussed in section 2.4, the framework of local feature detection, description, and matching is the most promising one. In chapter 4, we propose a feasible scheme to explore this framework and display its superiorities.

To achieve a robust recognition system, facial surface uniform resampling technique has been utilized in many successful approaches, such as the annotated deformable model based deformation modeling approaches [Kakadiaris *et al.* 2007a] [Passalis *et al.* 2011], the iso-level curve based approaches [Samir *et al.* 2006] [Drira *et al.* 2013], the local geometry-texture descriptor based approaches [Wang *et al.* 2010], and the local geometry-shape descriptor based approaches [Mian *et al.* 2008]. In chapter 5, as a similar technique,

surface remeshing with curvature convergence will be deeply discussed.

CHAPTER 3

3D Face Recognition in the Presence of Various Expressions

Contents

3.1 Introduction 67
3.1.1 Related work 68
3.1.2 Motivations and the proposed approach 69
3.2 Overview of the Proposed Approach 72
3.3 Local Normal Patterns (LNP) based Facial Descriptor 73
3.3.1 Facial Normal Estimation 73
3.3.2 Facial Normal Encoding 75
3.3.3 Facial Normal Representation 76
3.4 Weighted Sparse Representation-based Classifier 77
3.4.1 3D Face Subspace and Sparse Representation-based Classifier . . . 77
3.4.2 Weighted Sparse Representation-based Classifier 78
3.5 Experiments 80
3.5.1 Databases and Preprocessing 80
3.5.2 Experimental Settings 82
3.5.3 Experimental Results 84
3.6 Conclusion and Future Work 94

3.1 Introduction

Biometric systems are dedicated to identifying human beings from their own unique hard or soft physiological attributes such as the iris, face, fingerprint, hand vein, gait, gender, *etc*. Among these attributes, the face has proved to be one of the most popular and promising biometric modalities mainly due to the nature of human perception and the non-intrusiveness of face data acquisition. Although intensity image based 2D face recognition (FR) systems have provided solutions to achieve high performance under constrained conditions, the variations, especially caused by lighting and pose, which typically occur in an uncontrolled environment, are still its main obstacle [Zhao *et al.* 2003]. With the advent of 3D sensors, it is widely expected that 3D face scans, in providing geometrical information of facial surfaces, will open up a new avenue to handling these unsolved issues in 2D. As such, 3D face recognition (FR) has attracted increasing attention in recent years [Phillips *et al.* 2005, Bowyer *et al.* 2006].

3.1.1 Related work

A typical 3D FR algorithm comprises the following major components although they are strongly interwoven with each other [Gokberk *et al.* 2008]: 3D face landmarking, 3D face registration, the extraction of facial features along with the design of a matching scheme which closely depends upon the chosen facial features. Automatic 3D face landmarking is to automatically locate some key facial fiducial points, *e.g.*, the nose tip, inner eye corners, *etc.*, which are instrumental in face cropping, face alignment and pose normalization. The most challenging issue of automatic landmarking is to tolerate the disturbance caused by arbitrary variations in facial expression, pose, or occlusion [Zhao *et al.* 2011], and existing landmarking techniques are mainly based on the analysis of facial surface curvatures, shape index values, the facial symmetry central profile or depth information [Szeptycki *et al.* 2009a, Alyuz *et al.* 2010, Wang *et al.* 2010]. 3D face registration is aligning 3D face scans on a common coordinate system so that the matching of facial features can be carried out in a consistent way. Popular methods for the registration of 3D face scans are ICP-based [Faltemier *et al.* 2008a, Mian *et al.* 2007], consisting of minimizing the distance of two 3D point clouds in an iterative way, although they are reputed to be computationally expensive. The extraction of facial features means generating a discriminating facial representation which should comprehensively describe each 3D face scan for the latter stage of matching. As all human faces are similar to each other in terms of configuration whereas a 3D face scan accurately captures the geometrical shape of the underlying 3D facial surface, thereby making it likely to be more sensitive to facial expressions in comparison with 2D facial images, the design of a discriminating facial feature which remains robust to facial expressions is a critical issue in 3D FR. A number of approaches have been proposed in the literature, including facial curves [Samir *et al.* 2006], geometry and normal maps [Kakadiaris *et al.* 2007a], tensor based representations [Mian *et al.* 2008], iso-geodesic stripes [Berretti *et al.* 2010a], Multi-Scale Local Binary Pattern (MS-LBP) Depth maps and Shape Index (SI) maps [Huang *et al.* 2010], Multi-Scale extended Local Binary Pattern (MS-eLBP) maps [Huang *et al.* 2012], LBP-based normal angle encoding [Tang *et al.* 2013], *etc.* Other essays try to explicitly account for facial expression variations. An original tentative was made by Bronstein *et al.* [Bronstein *et al.* 2007] who assumed that facial expressions can be modeled as isometries of the facial surface and proposed a facial expression invariant canonical form. However, their assumption proves to be inexact, especially in the presence of exaggerated facial expressions [Drira *et al.* 2013]. A far more popular approach observes that facial expressions introduce facial distortions but there are still relatively stable facial regions, *e.g.*, forehead, nose region, from which expression robust features can be extracted [Ben Amor *et al.* 2008, Amor *et al.* 2006b]. Chang *et al.* [Chang *et al.* 2006] selected three regions around the nose for 3D face matching whereas Faltemier *et al.* [Faltemier *et al.* 2008a] extended this number to 28 small regions on the face. Tang *et al.* [Tang *et al.* 2013] proposed a 3D face specific feature-based sparse and dense division to capture the face structure and the face feature distribution. However, automatic detection and segmentation of facial surface into rigid and mimic regions is still problematic [Wang *et al.* 2010, Amor *et al.* 2006b].

The overwhelming majority of 3D FR algorithms proposed so far in the literature are evaluated on the FRGC v2.0 dataset [Phillips *et al.* 2005] which has de facto become the standard benchmark for 3D FR algorithms. Very high performance, up to 99% rank-

one recognition rate [Spreeuwers 2011], was reported on this dataset. However, although FRGC v2.0, with its 4007 3D face scans from 466 subjects, is the largest public 3D face dataset so far known in the literature, all its scans were captured in a frontal pose and in controlled lighting conditions, and less than half of them depict only a limited number of facial expressions, *e.g.*, happiness, surprise, *etc*. 3D face scans captured from uncooperative subjects in real-life applications can feature other challenges, *i.e.* missing data due to an arbitrary pose, external occlusions, and innumerable other types of facial expressions, whether subtle, prototypical or exaggerated. As a result, 3D FR algorithms with high performance on FRGC v2.0 can vastly degrade under other settings as revealed by the recent studies on the 3D Twins Expression Challenge (3D-TEC) database [Vijayan *et al.* 2011b, Vijayan *et al.* 2011a]. 3D-TEC stages a scenario of distinguishing 107 sets of identical twins through 3D face scans, each subject depicting a neutral and a smiling facial expression. This is a very challenging scenario for 3D FR systems because of the strong similarities between the 3D facial surfaces of twins, in addition to the traditional interference factors such as facial expression variations. Vijayan *et al.* [Vijayan *et al.* 2011b] evaluated the performance of four state-of-the-art 3D FR algorithms on the 3D-TEC dataset. They found that some algorithms performed very well on FRGC v2.0 but their performance greatly deteriorated on 3D-TEC, especially in the joint presence of disturbing factors, *i.e.* strong inter-class facial similarities and intra-class variations of facial expressions. Their results suggest that benchmarking 3D FR algorithms on FRGC v2.0 is certainly necessary but not sufficient to ensure the same performance and robustness with respect to the challenges of real-life applications, including in particular uncountable facial expressions which can lead to subtle, moderate and exaggerated facial surface deformations.

3.1.2 Motivations and the proposed approach

In this chapter, we target the challenge of facial expressions in 3D FR and propose a discriminative facial surface representation and an expression-robust method to handle expression variations. Specifically, we propose a novel facial shape descriptor, namely Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP), which represents the local facial shape by encoding their three normal components: x, y, and z respectively as binary patterns in a multi-scale way. An input facial surface can then be represented as a certain number of local normal pattern based maps or histograms of Local Normal Patterns (LNP).

As we know, surface curvatures [Alyuz *et al.* 2010, G.G.Gordon 1992, Sun *et al.* 2011] and shape index values [Huang *et al.* 2010, Zhang & Wang 2011, Li *et al.* 2011c] have been widely investigated for facial surface representation and characterization. However, the surface normal, which determines (at each point) the orientation of a facial surface, has not been fully explored in terms of 3D face representation [1]. Abate *et al.* [Abate *et al.* 2005, Abate *et al.* 2006, Abate *et al.* 2007a, Abate *et al.* 2009] introduced normal maps to describe facial surfaces, but this direct use of normal information in the holistic way did not achieve satisfying results. Gokbert *et al.* [Gokberk *et al.* 2006] used

[1] Note that, recently, normal constraint based surface registration for 3D FR methods such as [Queirolo *et al.* 2010, Mohammadzade & Hatzinakos 2013] have achieved very high performance on FRGC v2.0. In this paper, we focus on extracting a normal based facial descriptor for 3D FR.

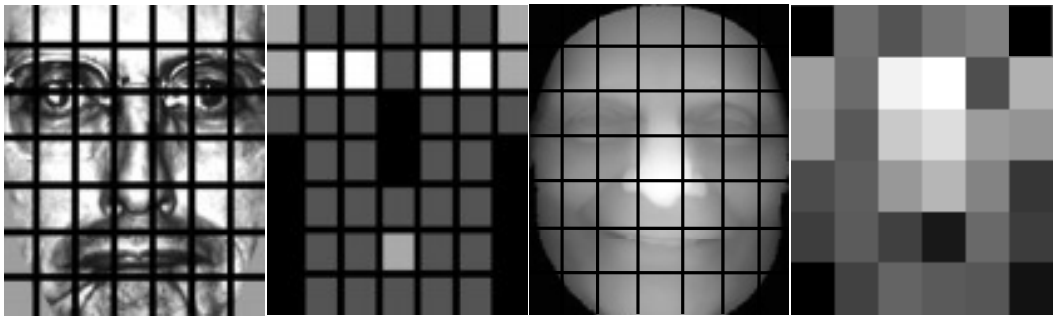
surface normal variance at each pixel location as a distance measure between face images and reported a rank-one score of 87.8% on the whole FRGC v2.0 database, while this performance greatly deteriorated on the 3D-TEC database [Vijayan *et al.* 2011b]. Kakadiaris *et al.* [Kakadiaris *et al.* 2007a] proposed to extract wavelet coefficients from normal and geometry maps for the computation of similarity, and reported a rank one recognition rate of 97% on the FRGC v2.0 database; however, the wavelet transform along with the fitting of the annotated deformable model is quite computationally expensive. Inspired by the competitive performance and computational efficiency of local binary patterns (LBP) for texture classification and 2D FR [Ojala *et al.* 2002, Ahonen *et al.* 2004a, Huang *et al.* 2011h], we propose to encode surface normal information, namely x, y, and z component normal images, in a local manner in order to generate histograms of LNP, in a way similar to that used by LBP for texture image description. The idea behind it lies in the fact that different facial shapes can be comprehensively described by the LNPs through different encoding scales and normal components, thereby making LNPs a very discriminative descriptor capable of recognizing 3D faces and even distinguishing between identical twins.

To pursue expression-robust 3D face recognition, we have seen that the popular method consists of choosing rigid facial regions, *e.g.*, the nose and forehead regions [Faltemier *et al.* 2008a, Chang *et al.* 2006], for the purpose of 3D face matching. However, the segmentation of a 3D facial surface into relatively rigid and elastic regions is problematic. Furthermore, such an approach also tends to ignore the elastic facial regions which also bear significant discriminating information. In this paper, we consider another alternative and propose to find the average quantification weights of meaningful local facial regions, *e.g.*, the eyes, nose, mouth, *etc.*, according to their discriminating power in 3D face recognition. This kind of quantification weights of local patches has been investigated in 2D FR and several works [Ahonen *et al.* 2004a, Martinez 2002, Lei *et al.* 2011] demonstrated its effectiveness (see Fig. 3.1 (a) and (b) for an example). It is interesting to note that to the best of our knowledge, the corresponding strategy in 3D using the weighting of local facial regions has not been studied yet for the purpose of expression-robust 3D FR . As shown in Fig. 3.1 (c) and (d), in this paper, we will show that the weights associated with 3D local facial regions are largely different from those of their 2D counterparts, especially in the regions of the nose and mouth. These weights can be learned from a given training set in the training phase (see Fig. G.1). The learned patch weights then can be used to build a weighted sparse representation model and compute the weighted reconstruction errors, leading to a Weighted Sparse Representation-based Classifier (W-SRC).

The main contributions of this chapter can be summarized as follows:

1. We introduce a new 3D facial representation based on Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP) for facial shape description. MSMC-LNP describes the micro-structure of facial normal information in multiple scales and multiple normal component channels. The extensive experiments show that the proposed LNP based facial representation is more discriminative than both the raw normal information and the encoded range image, *i.e.*, Local Shape binary Patterns (LSP) [Huang *et al.* 2011g] (see Tab. 3.3). We also show that the fusion of both multiple scales and multiple normal components is a helpful way of improving the final performance and we demonstrate its competency on 3D face identification as well as in the challenging issue of recognizing expressive identical twins.

2. A learning-based strategy is proposed to find out the quantification weights of local



(a) (b) (c) (d)

Figure 3.1: *Illustration of patch weights for 2D and 3D face recognitions: (a-b) a 2D face image and its corresponding patch weights [Ahonen et al. 2004a]; (c-d) a 3D face depth image and its corresponding patch weights learned by our method. All images are split to* 6×6 *local patches. Darker patches indicate lower weights, while brighter ones indicate higher weights.*

patches of 3D facial surfaces in terms of discriminating power. Given a training database, the patch weights associated with different facial regions, encoding scales, and normal components are learned by normalizing the patch matching scores. Those scores are computed by running the sparse representation-based classifier (SRC) over MSMC-LNP features of local patches. The experimental results show that the weights associated with these local patches in 3D are quite different from those of their 2D counterparts, especially in the nose region, thereby highlighting the fact that 3D and 2D face data bear different facial information which can be further explored in a multimodal face recognition scenario.

3. Using the learned patch weights along with the proposed MSMC-LNP feature, a weighted sparse representation-based classifier (W-SRC) is formulated to account for the different sensitivities of local facial regions to facial expressions, making the proposed 3D FR algorithm both discriminating enough to distinguish between identical twins and robust enough to tolerate various facial expressions. In particular, we formally establish that a weighted combination of SRCs amounts to directly weighting the corresponding feature vectors of those SRCs.

4. Extensive experiments were carried out using various 3D face datasets, including FRGC v2.0, BU-3DFE, Bosphorus and 3D-TEC, to benchmark the effectiveness and the generalizability of the proposed approach with respect to 3D face scans captured in different scenarios and conditions with different 3D sensors, depicting in particular different challenges in terms of facial expressions. This is in clear contrast with the overwhelming majority of 3D FR algorithms so far proposed in literature which are evaluated only on FRGC v2.0 with 3D face scans depicting a limited number of facial expressions. The experimental results demonstrate the effectiveness and the generalization ability of the proposed approach which consistently displays competitive rank-1 recognition rates over those datasets of different natures.

This chapter integrates our preliminary work in [Li *et al.* 2011a] but significantly extends that work. First, we carefully reformulate here the weighted sparse representation-based classifier (W-SRC) and formally establish that a weighted combination of SRCs amounts to directly weighting the feature vectors of those SRCs. Then, we highlight the

contributions of each design decision, *e.g.*, local normal patterns, the use of multi-scale encoding and normal components, the SRC-based learning strategy, with respect to the improvement of the recognition performance, and carry out an analysis of time complexity of the proposed 3D FR algorithm in comparison with the state of the art. Finally, we extensively evaluate the robustness of the proposed system with respect to a rich set of facial expressions, using in particular the BU-3DFE database which includes expressive 3D face scans depicting the six prototypical expressions in different intensities, as well as the Bosphorus database which features both the six prototypical expressions and subtle facial expressions through action units. This is in clear contrast with the overwhelming majority of 3D FR algorithms which were only benchmarked over FRGC v2.0 depicting merely a limited number of facial expressions. Besides, we validate the discriminative power of the proposed feature and the robustness of the proposed approach to facial expressions in the challenging issue of distinguishing between expressive identical twins using the 3D-TEC database. In addition, we also evaluate the computational cost of the proposed method.

The reminder of this chapter is organized as follows. The framework overview of the proposed system is presented in Section 2. Section 3 introduces the proposed Local Normal Patterns (LNP) based facial descriptor. Section 4 describes the weighted sparse representation-based classifier. In section 5, we present the experimental and algorithmic settings and discuss the experimental results. Section 6 concludes this chapter.

3.2 Overview of the Proposed Approach

As illustrated in Fig. G.1, the framework of the proposed approach consists of two phases: *i.e.* a training phase and a testing phase. Before the training and testing phases, each raw 3D face scan is preprocessed, *e.g.*, spike and noise removing, holes filling, nose tip localization, face cropping and alignment, to generate a range image with a predefined size. The training process is carried out in order to learn from a predefined training set the quantitative weights of local facial regions in terms of discriminating power and robustness in facial expression variations. It includes three procedures: feature extraction, identification and score normalization within different patches. They are briefly summarized as follows:

(1) Patch feature extraction. This procedure consists of three steps: (a) facial normal estimation; (b) facial normal encoding; (c) facial normal representation. Specifically, given a raw 3D face scan, we first launch the preprocessing pipeline (see Sec. 3.5.2) to normalize the range image to an $m \times n \times 3$ matrix (*i.e.*, x, y and z coordinates). Based on the range image, we estimate its three normal components (x, y, and z) by the local plane fitting method (see Sec. 3.3.1). Each normal component map is coarsely split into several local patches (*e.g.* 3×3); then each of these local patches is encoded as LNP with multiple scales, giving birth to multi-scale and multi-component local normal patterns (MSMC-LNP) to comprehensively describe the shape of each patch.

(2) Patch-based identification. Given a patch, an encoding scale, and a normal component, the corresponding LNP is extracted and fed into the sparse representation-based classifier (SRC) to generate a rank-one recognition rate using the training set.

(3) Patch score normalization. The patch scores, *i.e.* rank-one recognition rates, of different encoding scales and normal components are further normalized as the corresponding patch weights. The importance of facial physical regions in terms of discriminating power

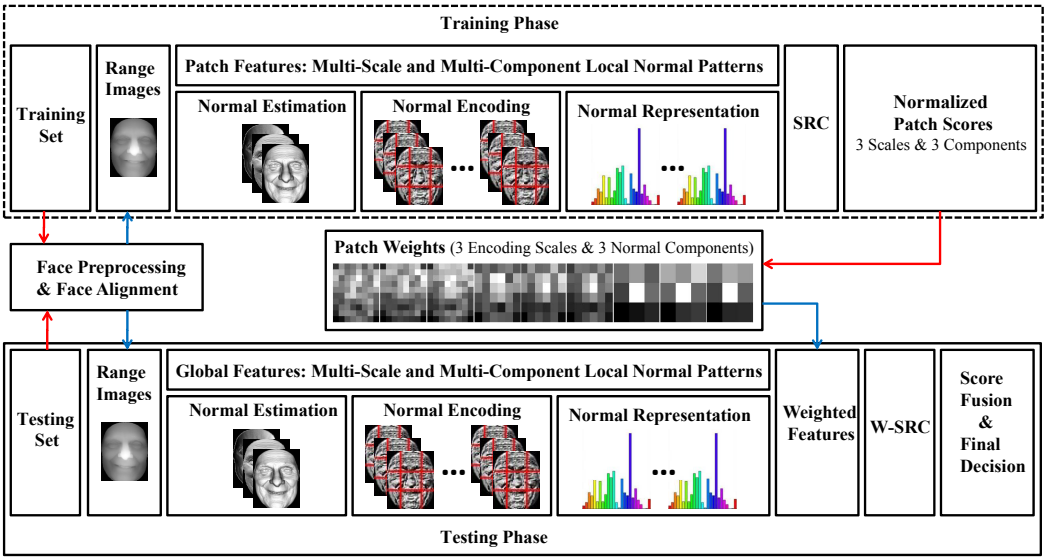


Figure 3.2: *Overview of the proposed approach.*

and robustness to facial expression variations can thus be measured by those quantitative patch weights.

During the testing phase, given a preprocessed range image in the testing set, we first compute the MSMC-LNP features over all the patches as in procedure (1) in the training phase. The global MSMC-LNP features are then obtained by simply stacking all these patch based MSMC-LNP features according to the holistic configuration of facial surfaces (see Sec. 3.3.3). Based on the patch weights learned in the training phase, weighted sparse representation (W-SRC) is formulated as seeking the sparse solution of the sum of the weighted patch based sparse representation (see (3.13) in Sec. 3.4). Then, W-SRC carries out face identification by finding the minimal weighted reconstruction residuals. (see (3.14) in Sec. 3.4). The final similarity measurement of MSMC-LNP in decision making is computed using the score level fusion which combines the matching scores of three encoding scales and three normal components through a simple sum rule.

3.3 Local Normal Patterns (LNP) based Facial Descriptor

3.3.1 Facial Normal Estimation

To highlight local variations of facial surfaces, remember that we make use of their normal information instead of the original point-cloud or range images. Existing normal estimation methods can be roughly classified into optimization based methods (*i.e.*, local fitting methods) and averaging methods [Klasing *et al.* 2009].

The basic idea of optimization based methods is as follows: 1) the normal vector on a given point can be calculated as the normal vector of a plane or the quadratic surface which it belongs to; 2) the underlying plane or surface can be estimated by fitting it to the local neighboring points around that given point; 3) the fitting problem then can be solved by minimizing a cost function penalizing a certain criterion, *e.g.*, the distance of the neighboring points from the searched local plane (see Fig. 3.3 (a)). The averaging methods

estimate the normal vector of a given point by computing a weighted average of the normal vectors of the triangles in its one-ring neighbors, the weights being the inverse ratios of the areas or the surrounding angles of the triangles in its one-ring neighbors (see Fig. 3.3 (b)).

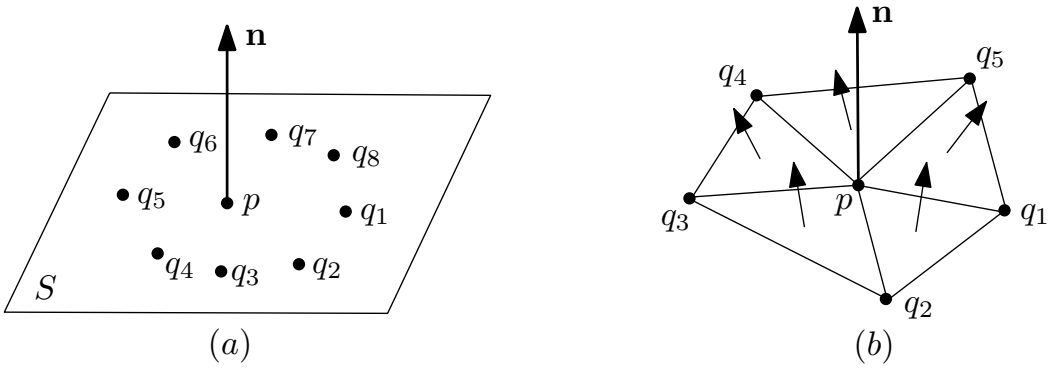


Figure 3.3: *Illustration of two approaches to normal estimation: (a) a plane is fitted to a vertex p and its neighbors; (b) the normal vectors of triangles in one-ring of p are averaged.*

The optimization-based methods can be applied to 3D point-clouds and triangular meshes while the averaging methods can only work on triangular meshes. Both types of methods are competent for normal calculation. In this paper, 3D face scans from various datasets were captured in different conditions and scenarios with divers 3D sensors, resulting in a diversity of data formats. As a result, an optimization-based method is adopted for the estimation of normals.

Specifically, given a facial range image **P** represented by an $m \times n \times 3$ matrix:

$$\mathbf{P} = [p_{ij}(x,y,z)]_{m\times n} = [p_{ijk}]_{m\times n\times\{x,y,z\}}, \tag{3.1}$$

where $p_{ij}(x,y,z) = (p_{ijx}, p_{ijy}, p_{ijz})^T, (1 \le i \le m, 1 \le j \le n, i, j \in \mathbb{Z})$ represents the 3D coordinates of the point p_{ij}. Let its unit normal vector matrix ($m \times n \times 3$) be

$$\mathbf{N}(\mathbf{P}) = [n(p_{ij}(x,y,z))]_{m\times n} = [n_{ijk}]_{m\times n\times\{x,y,z\}}, \tag{3.2}$$

where $n(p_{ij}(x,y,z)) = (n_{ijx}, n_{ijy}, n_{ijz})^T, (1 \le i \le m, 1 \le j \le n, i, j \in \mathbb{Z})$ denotes the unit normal vector of p_{ij}. As described in [Hoffman & Jain 1987], the normal vector **N(P)** of range image **P** can be estimated using the local plane fitting method. That is to say, for each point $p_{ij} \in \mathbf{P}$, its normal vector $n(p_{ij})$ can be estimated as the normal vector of the following local fitted plane:

$$S_{ij} : n_{ijx}q_{ijx} + n_{ijy}q_{ijy} + n_{ijz}q_{ijz} = d, \tag{3.3}$$

where $(q_{ijx}, q_{ijy}, q_{ijz})^T$ represents any point within the local neighborhood of point p_{ij} and $d = n_{ijx}p_{ijx} + n_{ijy}p_{ijy} + n_{ijz}p_{ijz}$. In this work, a neighborhood of 5×5 window is used. To simplify, each normal component in equation (2) can be represented by an $m \times n$ matrix:

$$\mathbf{N}(\mathbf{P}) = \begin{cases} \mathbf{N}(\mathbf{X}) = [n_{ij}^x]_{m\times n}, \\ \mathbf{N}(\mathbf{Y}) = [n_{ij}^y]_{m\times n}, \\ \mathbf{N}(\mathbf{Z}) = [n_{ij}^z]_{m\times n}. \end{cases} \tag{3.4}$$

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

Fig. 3.4 shows a sample range image extracted from the 3D-TEC database and its estimated three normal component matrices (images). As we can see from this figure, the normal component images contain more informative geometric information than their corresponding range image which is quite smooth. In particular, it can be seen that the geometric shape details around the eyes, mouth and forehead regions are quite well highlighted.

(a) (b) (c) (d)

Figure 3.4: *Illustration of facial normal estimation: (a) the original range image, (b-d) its normal images of component x, y and z (the sample comes from the 3D-TEC dataset).*

3.3.2 Facial Normal Encoding

Inspired by the discriminative power and computational simplicity of LBP for 2D texture description, we propose encoding each normal component, x, y, and z, respectively, as local normal patterns (LNP) to further highlight local shape variations. Thanks to the matrix form of these normal components as in equation (G.4), the encoding of these local shape variations can be carried out in a similar way to that used by LBP to encode 2D texture images. Specifically, the value of each point in a normal component is compared with its neighbors in a pre-defined neighborhood. A local neighborhood is defined as a set of sampling points evenly over a circle which is centered at the pixel to be labeled, and the sampling points that do not fall within the pixels are interpolated using bilinear interpolation, thus allowing for any radius and any number of sampling points in the neighborhood. Fig. 3.5 shows two examples of neighborhood of LNP, where the notation $Q_{r,p}$ denotes a neighborhood of p sampling points on a circle of radius of r.

After subtracting the central pixel value, the resulting strictly negative values are encoded with 0 and the others with 1; a binary number is thus obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one and its corresponding decimal value is used for labeling. The decimal numbers which result from such a process are referred to as local normal patterns (LNP). Formally, given a point p_{ij}, its normal component noted as $n_{ij}^k(0)$, the derived LNP decimal value is:

$$LNP(Q_{r,p}(p_{ij})) = \sum_{q=1}^{p-1} t(n_{ij}^k(q) - n_{ij}^k(0))2^q, \tag{3.5}$$

where $t(x) = 1, \text{if } x \geq 0$ and $t(x) = 0, \text{if } x < 0$.

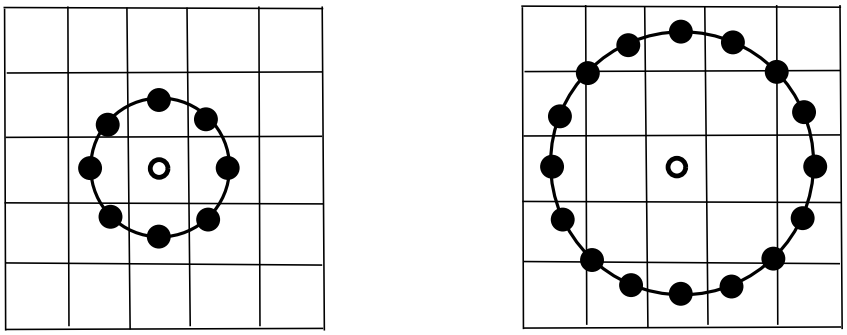


Figure 3.5: *Examples of the neighborhood of LNP: (a) $Q_{1,8}$ and (b) $Q_{2,16}$.*

Given a normal component, $LNP(Q_{r,p})$ thus encodes local normal variations around each point as a decimal value, denoted as $e([n_{ij}^k]_{m\times n}), k \in \{x,y,z\}$. Fig. 3.6 gives an example of $LNP(Q_{1,8})$ on the three facial normal components of the same subject.

Figure 3.6: *Illustration of facial normal encoding: (a) to (c), normal images of component x, y and z; (d) to (e), their corresponding LNP maps using the neighborhood $Q_{1,8}$.*

LNP encodes the local shape variations at point level. In order to characterize the shape of a local region, histogram-based statistics are computed and used as a facial feature vector. For a given normal component $k \in \{x,y,z\}$, the histogram of the encoded normal component $e([n_{ij}^k]_{m\times n})$ can be defined as:

$$H = \sum_{i,j} I\{e([n_{ij}^k]_{m\times n}) = s\}, s = 0,\ldots,S-1, \tag{3.6}$$

where s is the encoded decimal number, and $S-1$ is the maximum value of local normal patterns (LNPs) given a neighborhood $Q_{r,p}$, *e.g.*, for $Q_{1,8}, S = 2^8 = 256$. $I\{A\} = 1$, if A is true, else $I\{A\} = 0$. This histogram describes the local micro-patterns of a given normal component over the whole face model.

3.3.3 Facial Normal Representation

To utilize the spatial information of 3D face scans, each facial normal component, x, y, and z, can be further divided into several patches, from which LNP histograms H are extracted

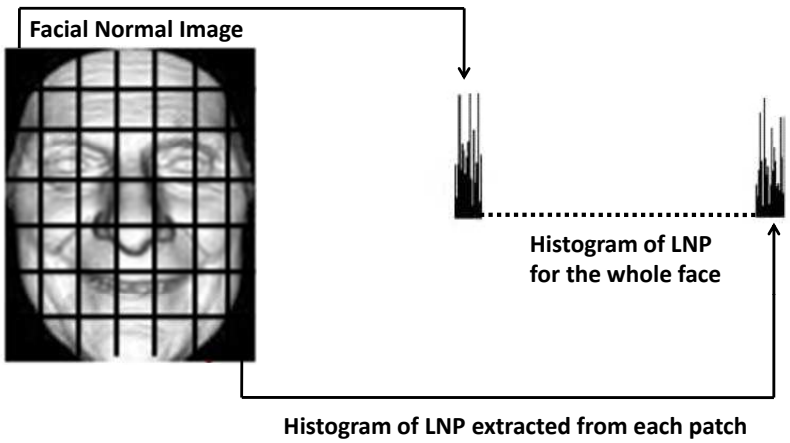


Figure 3.7: *Illustration of facial normal representation: histogram of LNP.*

and then concatenated to form a global histogram G which thus captures the facial configuration of the encoded facial normal feature (see Fig. 3.7). Finally, the original facial surface is described by three global feature histograms G_x, G_y, and G_z on a given encoding scale.

3.4 Weighted Sparse Representation-based Classifier

Once 3D face scans have been described by their respective facial normal representations, a proper classifier is needed for the purpose of 3D face identification. In this work, we introduce a Weighted Sparse Representation-based Classifier (W-SRC) which is described in this section. We first introduce the 3D face subspace based sparse representation model and its corresponding SRC. Then, we theoretically formulate a weighted sparse representation model and its corresponding W-SRC to account for the patch weights which are learned using a training set to quantify the relative discriminating power of each local facial region. We demonstrate in particular that solving a weighted SRC amounts to computing a single SRC with global feature vectors in stacking the weighted feature vectors of each patch.

3.4.1 3D Face Subspace and Sparse Representation-based Classifier

A common assumption in 2D FR is the 2D face subspace model which assumes that well-aligned frontal face images in different lighting conditions and various facial expressions, approximately lie on a special low-dimensional linear subspace spanned by sufficient training samples from the same subject. Using such an assumption, Wright *et al.* [**?**] proposed a sparse representation model and its corresponding SRC for robust 2D FR. In this work, we make an extension of this assumption to the 3D case, and assume that well-aligned frontal 3D face scans with different facial expressions approximately lie on a special low-dimensional linear subspace spanned by sufficient training 3D face scans from the same subject. We call this assumption a 3D face subspace model. Formally, it can be formulated by the following equation:

$$y \approx \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_N v_N. \tag{3.7}$$

That is, given N_i training samples of the i-th subject, $[v_{i,1}, v_{i,2}, \ldots, v_{i,N_i}] \in \mathbb{R}^{M \times N_i}$, any test sample $y_i \in \mathbb{R}^M$ of the i-th subject can be represented, according to (3.7), as:

$$y_i \approx \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + ... + \alpha_{i,N_i} v_{i,N_i}, \tag{3.8}$$

where $\alpha_{i,j} \in \mathbb{R}, j = 1, 2, \ldots, N_i$.

However, it is worth noting that there is only one training sample for each subject in the gallery in the most common experimental setting in 3D FR. Without occlusion, the only difference between two well-aligned frontal 3D face scans of the same subject is the local shape distortion caused by variations in expression. This problem of insufficient training samples along with the shape distortion caused by expression variations introduces a new model error term, denoted as $\varepsilon_i \in \mathbb{R}^M$. Thus, model (3.8) can be modified as:

$$y_i \approx \alpha_{i,1} v_{i,1} = \alpha_{i,1} v_{i,1} + \varepsilon_i, \tag{3.9}$$

where $y_i \in \mathbb{R}^M$, $v_{i,1} \in \mathbb{R}^M$ and $\alpha_{i,1} \in \mathbb{R}$ represent a probe face, a gallery face from the same subject and their linear scalar factor respectively. Using (3.9), sparse representation model and its corresponding SRC for 3D FR can be formulated as follows:

Given a gallery set with N 3D face scans, each of which belongs to one subject, we define the dictionary as $D \doteq [v_1, v_2, \ldots, v_N] \in \mathbb{R}^{M \times N}$. Then for any probe $y \in \mathbb{R}^M$, we have

$$y = Dx + \varepsilon, \tag{3.10}$$

where $x = [0, \ldots, 0, \alpha_j, 0, \ldots, 0]^T \in \mathbb{R}^N$ is the coefficient vector whose entries are zero except the one associated with the j-subject. Sparse coefficients x in (G.13) can be solved by the following l_0 minimization problem:

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \|y - Dx\|_2^2 \leq T, \tag{3.11}$$

where $T = \|\varepsilon\|_2^2$.

In practice, we employ the Orthogonal Matching Pursuit (OMP) [Pati *et al.* 1993] algorithm to solve (3.11) and compute the reconstruction residuals:

$$r_i(y) = \|y - D\delta_i(\hat{x})\|_2^2, i = 1, 2, \ldots, N. \tag{3.12}$$

where δ_i is a characteristic function which selects the coefficient associated with the i-th gallery. Finally, the index of minimal $r_i(y)$ delivers the identity of the probe y.

3.4.2 Weighted Sparse Representation-based Classifier

Assume now that each face scan is divided into L different patches. Denote w_l as the learned weight for patch l. Using the MATLAB convention:

$$[x_1; x_2] \doteq \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

the feature vector v_i can be rewritten as

$$v_i = [v_{i1}; v_{i2} \dots v_{iL}],$$

where $v_{il} \in \mathbb{R}^{(M/L)\times 1}$, and the dictionary D can be denoted as

$$D = [D_1; D_2 \dots D_L],$$

where $D_l = [v_{1,l}, v_{2,l}, \dots, v_{i,l}, \dots, v_{N,l}]$, and a probe y can be denoted as

$$y = [y_1; y_2 \dots y_L],$$

where $y_l \in \mathbb{R}^{(M/L)\times 1}, l = 1, 2, \dots, L$.

Eq. (3.11) can then be rewritten as the following weighted sparse representation model:

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \sum_{l=1}^{L} w_l \|y_l - D_l x\|_2^2 \leq T, \tag{3.13}$$

and the corresponding weighted reconstruction residuals are

$$r_i(y) = \sum_{l=1}^{L} w_l \|y_l - D_l \delta_i(\hat{x})\|_2^2, i = 1, 2, \dots, N. \tag{3.14}$$

To solve eq. (3.13), we notice that it equals to solve

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \sum_{l=1}^{L} \|w_l y_l - w_l D_l x\|_2^2 \leq T. \tag{3.15}$$

We write

$$W(D) = [w_1 D_1; w_2 D_2 \dots w_L D_L],$$

and

$$W(y) = [w_1 y_1; w_2 y_2 \dots w_L y_L].$$

Then Eq. (3.15) is equal to

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \|W(y) - W(D)x\|_2^2 \leq T. \tag{3.16}$$

Eq. (3.16) means that the weighted sparse representation model as expressed in Eq. (3.13) amounts to solving a single SRC with global feature vectors by simply stacking weighted features of the corresponding patches. Eq. (3.16) can be solved by the OMP algorithm [Pati *et al.* 1993]. Once determined the sparse representation coefficient $\hat{x}$ of Eq. (3.16), weighted reconstruction residuals in Eq. (3.14) can be computed. Then the minimal $r_i(y)$ can be used to determine the identity of y. We call this sparse representation-based classifier enhanced by spatial weights the Weighted Sparse Representation-based Classifier (W-SRC) in the subsequent.

3.5 Experiments

Recent studies [Vijayan *et al.* 2011b] show that 3D FR algorithms with very high performance on the de facto standard benchmark, namely FRGC v2.0, can vastly deteriorate when evaluated on expressive 3D face scans captured in a different scenario. As a result, we decide to evaluate the effectiveness and the generalizability of the proposed approach using various datasets, namely FRGC, BU-3DFE, Bosphorus and 3D-TEC, with 3D face scans depicting a rich set of facial expressions, subtle, prototypical and exaggerated. Depending on the underlying dataset, these expressive 3D face scans were captured in different scenarios and conditions, *e.g.*, 3D sensors, lighting conditions, and thus offer different levels of challenge with respect to facial expression variations. In this section, we first introduce the different datasets and describe their particularity, then present the different experimental settings in order to highlight the various facets of the proposed approach and finally discuss the experimental results.

3.5.1 Databases and Preprocessing

In our experiments, three databases, namely FRGC v1.0 [Phillips *et al.* 2005], BU-3DFE [Yin *et al.* 2006a] and Bosphorus [Savran *et al.* 2008], are used as training sets to learn the the respective patch weights, while four databases, the BU-3DFE, Bosphorus, FRGC v2.0 [Phillips *et al.* 2005] and 3D-TEC [Vijayan *et al.* 2011a] are used as testing sets for cross database validation and evaluation. Fig. 3.8 plots a raw 3D face scan from each of these databases. They are briefly introduced as follows:

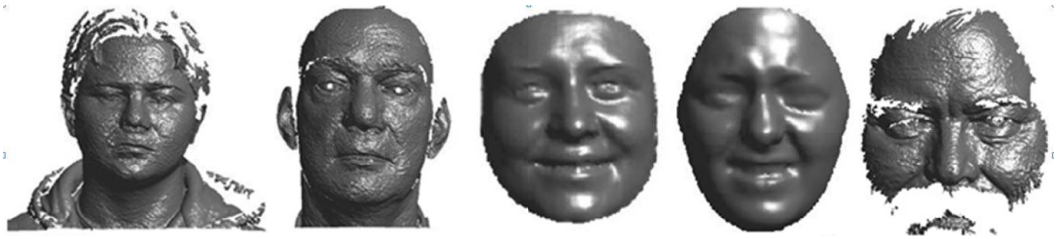


Figure 3.8: *Illustrate of the raw samples of the five databases: (a) FRGC v1.0, (b) FRGC v2.0, (c) Bosphorus, (d) BU-3DFE, (e) 3D-TEC.*

- ***FRGC v1.0*:** The FRGC v1.0 database (Spring2003) consists of 943 textured 3D face models of 275 subjects with a neutral expression. The hardware used to acquire these range images is a Minolta Vivid 900 (MV 900) laser range scanner, with a resolution of 640×480.

- ***FRGC v2.0*:** The FRGC v2.0 database (Fall2003 and Spring2004) is made up of 4007 textured 3D face models of 466 subjects with different facial expressions. The same hardware as in FRGCv1.0 is used for data acquisition, and the resolution of each range image is also 640×480. FRGC v2.0 is the largest public 3D face dataset which includes 1642 expressive frontal 3D face scans captured in controlled lighting conditions. *Although FRGC v2.0 has become de facto the standard benchmark for*

evaluating 3D FR algorithms, in particular because of its size, it contains only a very limited range of facial expressions, including happiness, surprise, disgust, sadness and so on. This is to be compared with the uncountable number of facial expressions which can occur in real-life applications in uncontrolled conditions. As a result, a 3D FR algorithm with high performance on FRGC v2.0 does not guarantee the same level of performance on other 3D face scans captured in different conditions and scenarios. The generalizability of a claimed robustness to facial expression variations needs to be checked on representative expressive 3D face scans.

- ***BU-3DFE*:** The BU-3DFE database contains 100 subjects (56 females and 44 males), ranging in age from 18 to 70 years old, with a variety of ethnic origins. Each subject performs seven expressions. Except for neutral, each of the six prototypic expressions (happiness, disgust, fear, anger, surprise, and sadness) includes four levels of intensity. Therefore, there are 25 instant models for each subject, resulting in a total of 2,500 3D facial models. The 3D models are captured with a 3D face imaging system named a 3DMD digitizer. Each model is saved as a polygonal mesh with a resolution ranging from 20,000 to 35,000 polygons. *This dataset thus makes it possible to benchmark 3D FR algorithms over 3D face scans depicting prototypical facial expressions of different intensities, ranging from very subtle to exaggerated, and quantitatively measures the robustness of a given 3D FR technique when varying the intensities of facial expressions.*

- ***Bosphorus*:** The Bosphorus database contains 4666 textured 3D face models of 105 subjects in various facial expressions, action units, poses and occlusions. The 3D models are acquired with a device named the Inspeck Mega Capturor II (IMC II). Each model is saved as a range image with a resolution of $1,600 \times 1,200$. *This dataset thus allows benchmarking 3D FR techniques in an uncontrolled environment in nearly real-life conditions. In this work, we are interested in Bosphorus for the difficulty it presents with facial expressions which contain not only prototypical expressions but also action units.*

- ***3D-TEC*:** The 3D-TEC database consists of 106 pairs of identical twins and a set of triplets, totalizing 214 subjects. Each subject contains two scans: one neutral scan and one smiling scan. More details can be found in [Vijayan *et al.* 2011b]. *This dataset thus enables the evaluation of 3D FR algorithms in difficult conditions, i.e. the joint presence of strong inter-class similarities and intra-class expression variations.*

All scans of FRGC v1.0, FRGC v2.0, and 3D-TEC databases are preprocessed by using the *3D Face Models Preprocessing Tool* [2] developed by Szeptycki *et al.* [Szeptycki *et al.* 2009a]. The preprocessing pipeline contains: spike and noise removing, holes filling, nose tip localization and face cropping. As introduced in [Szeptycki *et al.* 2009a], a decision-based median filtering technique is used to remove spikes, and the holes are detected by searching vertexes having less than 8 neighbors, and filled by fitting square surfaces. The nose tip is located through a curvature analysis-based

[2] http://pszeptycki.com/tool.html

coarse grained search and generic face model-based fine grained search. Exceptionally for 3D-TEC, the manually labeled nose tips provided by the database are used in this work. Finally, each scan is cropped by a sphere centering at the nose tip and with a radius of 90 mm. The polygon surface scans in BU-3DFE are first preprocessed as discrete manifold triangular meshes and then projected as range images by an interpolation algorithm. Then, nose tip localization and face cropping are carried out for all the scans of BU-3DFE and Bosphrous databases by using the same aforementioned preprocessing tool. Regarding the registration, we select a face scan with a neutral expression in frontal pose for each of the five databases as a reference model, and all the other face scans are then aligned to the reference model using the Iterative Closest Point (ICP) [Zhang 1994] algorithm. Fig. 3.9 illustrates several examples of preprocessed face models.

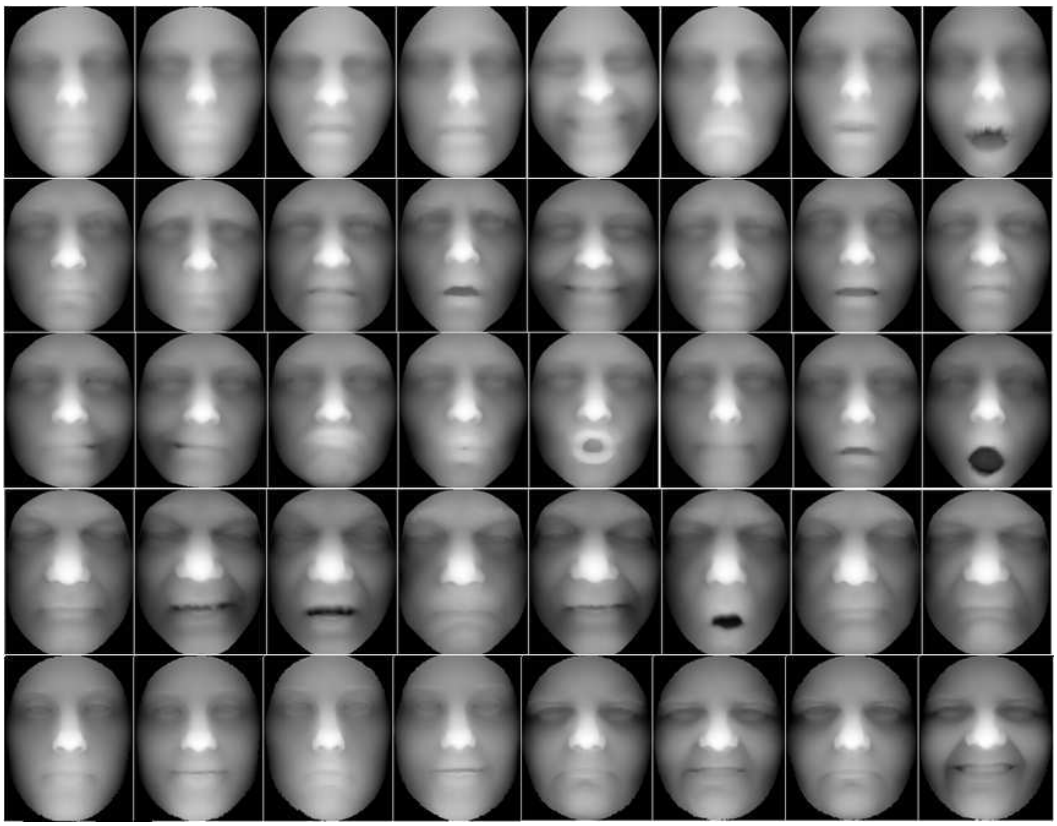


Figure 3.9: *Illustration of several examples of preprocessed face models: first row: models of one subject with different facial expressions (BU-3DFE); second and third rows: models of one subject with different facial expressions and action units (Bosphorus); fourth row: models of one subject with different facial expressions (FRGC v2.0); last row: two pairs of identical twin models with neutral and smiling expressions (3D-TEC).*

3.5.2 Experimental Settings

To comprehensively evaluate the proposed approach and highlight its various facets, seven experiments were designed. 1) The discriminative power of the proposed LNP descrip-

tor. This experiment aims to compare the proposed LNP with the raw normal maps and the direct encoding of range images to highlight the effectiveness of the proposed encoding to normal maps; 2) The effectiveness of SRC. This experiment aims to highlight the advantage of using a sparse representation-based classifier (SRC) in comparison with the popular Chi-square distance when histogram-based features (*e.g.*, LNP) are involved; 3) The impact of multi-scale encoding of LNPs and the benefit of using multi-component of normal images; 4) The patch weights learning and the effectiveness of W-SRC. This experiment is to show the learning process of patch weights, to bring out the difference of patch weights between 3D and 2D face data, and to highlight the effectiveness of the proposed W-SRC in terms of performance improvement; 5) The robustness analysis of the proposed approach with respect to various facial expression variations. This experiment aims to highlight the behavior of the proposed approach with respect to facial expressions as well as their intensities; 6) The performance of distinguishing between expressive identical twins. This experiment aims through 3D-TEC to benchmark the proposed 3D algorithm in difficult conditions, *i.e.* the recognition of expressive identical twins with strong inter-class similarities in presence of intra-class variations introduced by facial expressions. 7) The comparison with the state-of-the-art. This study compares the proposed approach with the state of the art using the four datasets whenever state of the art results are available.

The experimental settings are as follows: for the FRGC v1.0 and FRGC v2.0 databases, the first scan of each subject is used to make a gallery set and the remaining 3D face scans are treated as probes. As a result, 571 scans of 267 subjects from FRGC v1.0 and 3541 scans of 466 subjects from FRGC v2.0 are used as probes; for the BU-3DFE database, the neutral scan of each subject is used to make a gallery set and the remaining scans are treated as the probe set. Therefore, 2400 scans of 100 subjects are used as probes. For the Bosphorus database, we focus on the challenge of facial expression variations and select the first neutral scan to make the gallery set whereas the remaining scans with frontal pose and without occlusions are treated as probes, thus excluding those scans with pose variations and occlusions which depict different challenges. This results in a total of 2797 scans of 105 subjects being used as probes. Table 3.1 summarizes these protocols.

Table 3.1: *Experimental settings of FRGC v1.0, BU-3DFE, Bosphorus, and FRGC v2.0 databases.*

Database	**Subject**	**Scan**	**Gallery**	**Probe**
FRGC v1.0	267	838	267	571
FRGC v2.0	466	4007	466	3541
BU-3DFE	100	2500	100	2400
Bosphorus	105	2902	105	2797

The gallery and probe scans used for 3D-TEC database are based on the standard protocol shown in Table 3.2 [Vijayan *et al.* 2011b]. One person in each pair of twins is arbitrarily labeled Twin A and the other Twin B, and four Cases are considered. In Case I, all the images in the gallery set possess a smiling expression while all the images in the probe set have a neutral expression. Case II reverses Case I roles. In Case III, Twin A smiling and Twin B neutral make up the gallery set; while Twin A neutral and Twin B smiling as probe compose the probe set. Case IV reverses these roles of Case III. As pointed out in

[Vijayan *et al.* 2011b], theoretically the main challenge would be to distinguish between the probe image and the image of his/her twin in the gallery. Case III and IV are more difficult than Cases I and II since the expression of the probe face is different from his/her image in the gallery but the same as the image of his/her twin in the gallery. For all four cases, both gallery and probe sets consist of 214 scans of 214 subjects.

Table 3.2: *Experimental setting of 3D-TEC database: "A Smile, B Neutral" means that the set contains all images with Twin A smiling and Twin B neutral [Vijayan et al. 2011b].*

No.	Subject	Scan	Gallery	Probe
I	214	428	A Smile, B Smile	A Neutral, B Neutral
II	214	428	A Neutral, B Neutral	A Smile, B Smile
III	214	428	A Smile, B Neutral	A Neutral, B Smile
IV	214	428	A Neutral, B Smile	A Smile, B Neutral

Each preprocessed range image is further resized into 120×96 for facial normal estimation, resulting in three normal component matrices or images in sizes of 120×96. Each normal component matrix is divided into 10×8, 6×6 and 3×3 windows corresponding to local patches with sizes of 12×12, 20×16 and 40×32, respectively. Then, three different scales are considered with all kinds of local patches. This amounts to performing encoding operators $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$ on local patches with sizes of 12×12, 20×16 and 40×32, respectively. As a result, for each normal component, we encode it with three different scales, achieving three histograms of local normal patterns (LNP). In order to reduce the dimensionality of final facial features, the uniform pattern strategy [Ojala *et al.* 2002] as in LBP is adopted to decrease the number of bins in each local patch. Finally, from one original 3D face scan, we generate 9 histograms of local normal patterns (3 normal components and 3 encoding scales) involving both local patch based and global features. Each histogram representation of the whole face is fed into the classifier to achieve one similarity score matrix. All the 9 similarity score matrices are then fused through a simple sum rule to compute the final accuracy of MSMC-LNP. To solve (3.11) and (3.16), the Orthogonal Matching Pursuit (OMP) [Pati *et al.* 1993] algorithm with the sparse number of 30 for x is used in all the experiments.

3.5.3 Experimental Results

3.5.3.1 Experiment I: The discriminative power of local normal patterns

To highlight the discriminating power of the proposed LNP based facial feature, we compare it with other two kinds of facial features: i) The raw normal information based facial features Nx, Ny and Nz, by simply stacking the columns of each normal component matrices n_{ijx}, n_{ijy} and n_{ijz} respectively, and their fusion Nxyz. ii) Local Shape binary Patterns (LSP), *i.e.* LBP histograms extracted directly from range images. For a fair comparison, LNP descriptor used the same encoding parameter (*i.e.* $Q_{2,16}$) with LSP to extract the feature vector on each normal component, respectively noted as LNPx, LNPy and LNPz, and their fusion, *i.e.* Multi-Component Local Normal Patterns (MC-LNP). All features were finally fed into a SRC classifier. Once again, the score-level fusion through a simple sum

rule was employed to combine different normal components and encoding scales.

Table 3.3: *Comparison of rank-one scores: original normal, LSP and LNP on the whole FRGC v2.0 database.*

Approaches	Rank-one Scores
(1) Nx + SRC	67.83%
(2) Ny + SRC	65.62%
(3) Nz + SRC	71.63%
(4) Nxyz + SRC	73.19%
(5) $LSP_{2,16}$ + SRC	82.07%
(6) LNPx($Q_{2,16}$) + SRC	**87.01%**
(7) LNPy($Q_{2,16}$) + SRC	**86.13%**
(8) LNPz($Q_{2,16}$) + SRC	**88.43%**
(9) MC-LNP($Q_{2,16}$) + SRC	**92.60%**

Table 3.3 reports the rank-one recognition rates on the whole FRGC v2.0 database. We can see that LNP performs much better (about 20% higher) than the original normal feature. On the other side, without normal information, the result based on LSP is about 5% lower than that of each encoded normal component and 10% lower than their fusion, *i.e.* MC-LNP($Q_{2,16}$). This experiment indicates that the encoded normal information (LNP) is more discriminative, not only more so than the original normal information, but also than the encoded depth information (LSP).

3.5.3.2 Experiment II: The effectiveness of sparse representation-based classifier

For histogram based feature vector (*e.g.*, LNP), Chi-Square distance is the preferred similarity measurement in the literature [Pele & Werman 2010]. Table 3.4 compares the rank-one recognition rates achieved by SRC and Chi-Square distance based classifiers on the whole FRGC v2.0 database. All the results are achieved using LNP feature with the same encoding scale (i.e. $Q_{2,16}$).

Table 3.4: *Comparison of rank-one scores: Chi-Square vs. SRC on the whole FRGC v2.0 database.*

Approaches	Rank-one Scores
(1) LNPx($Q_{2,16}$) + Chi-Square	77.36%
(2) LNPx($Q_{2,16}$) + SRC	**87.01%**
(3) LNPy($Q_{2,16}$) + Chi-Square	77.87%
(4) LNPy($Q_{2,16}$) + SRC	**86.13%**
(5) LNPz($Q_{2,16}$) + Chi-Square	81.33%
(6) LNPz($Q_{2,16}$) + SRC	**88.43%**
(7) MC-LNP($Q_{2,16}$) + Chi-Square	82.64%
(8) MC-LNP($Q_{2,16}$) + SRC	**92.60%**

As can be seen from Table 3.4, the rank-one scores of SRC using LNPx, LNPy and

LNPz as well as their fusion MC-LNP, with an average gain of 8 points, consistently outperform those of the Chi-square distance-based classifier using the same feature vectors. These results highlight the effectiveness of SRC when using local normal pattern (LNP) based facial representation.

3.5.3.3 Experiment III: The effectiveness of multi-scale and multi-component local normal patterns

As described in subsection 3.5.2 on the experimental settings, the proposed 3D FR method makes joint use of three normal images, along with three different encoding scales through 3 regular grids of different sizes. This results in a global descriptor, namely MSMC-LNP, which is made up of 9 histograms of local normal patterns (*i.e.*, 3 normal components and 3 encoding scales). One interesting question here is thus how discriminating each normal component is in the task of 3D FR with the individual encoding scheme through a specific regular grid overlaid on 3D face scans. In this subsection, we propose to study the discriminating power of each normal component on a given encoding scale and its complementarity for the 3D FR task through FRGC v2.0.

Specifically, in the proposed 3D FR method, three regular grids of different size, namely 10×8, 6×6, and 3×3, are overlaid on 3D face scans. For a range image that we normalize to the size of 120×96, these regular grids lead to patch sizes of 12×12, 20×16 and 40×32, respectively. They thus capture the spatial relationships of different granularity, from the finest grid of 10×8 with a patch size of 12×12 to the coarsest grid of 3×3 with a patch size of 40×32, the grid of 6×6 with a patch size of 20×16 being the intermediate size.

Associated with these three different grids and thus three different patch sizes, LNP of three different scales, namely $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$, are computed on each component of the normal images, leading to LNPx, LNPy and LNPz, respectively on three different scales. The LNP of the finest scale $Q_{1,8}$ is computed over the finest grid 10×8, the LNP of the coarsest scale $Q_{3,24}$ over the coarsest grid 3×3 and the LNP of the intermediate scale $Q_{2,16}$ over the intermediate grid 6×6. These LNPs capture the shape variations of the underlying facial surface on different scales, from the finest variations to the coarsest ones.

Table 3.5 lists the performance of each normal component, namely LNPx, LNPy and LNPz in rows, using three different encoding schemes, namely $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$ in columns, and tabulates the performance of their fusion using the three normal components (row MC-LNP) and the three encoding scales (column MS-LNP). The performance of the whole fusion using the three different normal components through three encoding scales is reported in the row MSMC-LNP. As we can see from this table, all the three normal components on the three different encoding scales depict a more or less similar rank-one recognition rate with LNPz being (just) the best. However, when fused together, their performance is greatly improved, with more than 8 points when the coarsest encoding scale is used.

Now when analyzed from the viewpoint of the encoding scale along with the grid size, we can find from Table 3.5 that the best encoding scale is $Q_{2,16}$, which corresponds to the grid of intermediate granularity 6×6 with a patch size of 20×16. This result suggests that, if only one encoding scale needs to be chosen for the purpose of 3D face

recognition, scale $Q_{2,16}$ is the best tradeoff as it makes it possible to capture enough details of a facial shape to efficiently discriminate between people while remaining sufficiently robust to facial expression variations. However, the encoding at different scales captures different shape information and its fusion in the column MS-LNP results in an improvement in their individual rank-one recognition rate which can attain more than 4 points in the case of LNPy.

Table 3.5: *The effectiveness of multi-scale and multi-component local normal patterns on the whole FRGC v2.0 database. The encoding scales, namely $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$, are computed over regular grid 10x8 with patch size 12x12, grid 6x6 with patch size 20x16, and grid 3x3 with patch size of 40x32, respectively. The range image size is normalized to 120x96 in this work. Fusion is carried out at score level through simple sum rule.*

	$Q_{1,8}$	$Q_{2,16}$	$Q_{3,24}$	MS-LNP
LNPx	86.41%	87.01%	75.52%	**89.05%**
LNPy	85.39%	86.13%	80.17%	**90.47%**
LNPz	87.69%	88.43%	80.31%	**90.89%**
MC-LNP	**90.78%**	**92.60%**	**88.74%**	
MSMC-LNP		**94.13%**		

As the three different normal components and encoding schemes capture different shape information, we also perform the fusion of these two types of information, *i.e.*, combining the three normal components at 3 different encoding scales. It can be seen from the row MSMC-LNP in Table 3.5 that this overall fusion, *i.e.*, MSMC-LNP, does further improve the recognition accuracy which now attains 94.13% which is 2 points higher than the best fusion result using the three normal components and roughly 4 points higher than the best result achieved by fusing the three different scales. However, all the previous results are achieved by brute force application of our MSMC local normal patterns and do not make use of training datasets which should enable the weighting of these facial patches based on their discriminating power and robustness to facial expressions. The following section depicts the effectiveness of such a learning strategy to further improve recognition accuracy.

3.5.3.4 Experiment IV: The patch weight learning and the effectiveness of W-SRC

In this experiment, we first describe the way to learn patch weights and analyze the relative importance of local facial regions for face identification. Then, we compare the performance of W-SRC and SRC on FRGC v2.0, Bosphorus, and BU-3DFE respectively. Three databases are used for learning the patch weights: FRGC v1.0, BU-3DFE, and Bosphorus. The experimental protocol listed in Table 3.1 is used, and according to the proposed framework, the patch weights are achieved by the following four steps: 1) divide each normal component into local patches (10×8, 6×6, and 3×3 windows); 2) extract patch based MSMC-LNP features, three normal components and three encoding scales ($Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$); 3) compute patch based rank one scores using the SRC classifier on a given

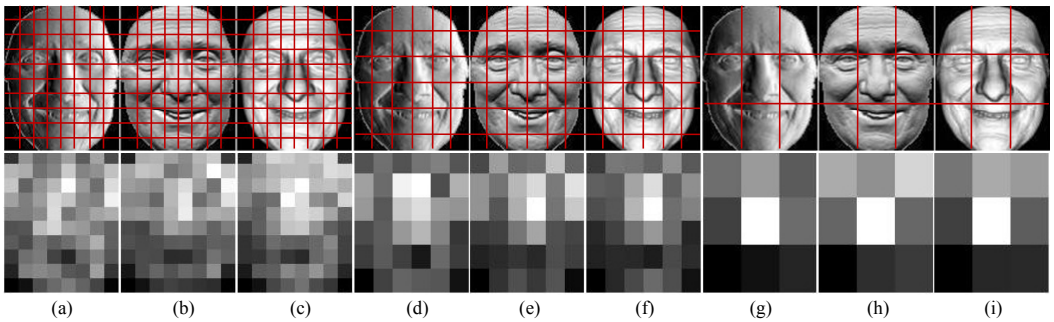


Figure 3.10: *Illustration of the patch weights learned from the Bosphorus database. Columns (a-c), normal images x, y and z and their patch weights (*10×8 *patches); columns (d-f), normal images x, y and z and their patch weights (*6×6 *patches); columns (g-i), normal images x, y and z and their patch weights (*3×3 *patches). Darker patches indicate lower weights, while brighter ones indicate higher weights.*

training database. 4) compute patch based weights by normalizing the patch based scores.

Fig. 3.10 shows the results of such a learning process using the Bosphorus as the training database in displaying the patch weights of three normal component images x, y and z with three binary encoding scales $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$. The patch numbers with respect to those three encoding scales are 10×8, 6×6 and 3×3, respectively. The weights are marked by gray values where darker ones indicate lower weights while the brighter ones indicate higher weights. We can see that the weight distribution patterns are quite different from each other among different normal components and different encoding scales but we can observe some similar trends, in particular with the largest weights near the nose regions, and larger weights near the eyes, while the smallest weights are found near the mouth regions and the boundary parts. For more detail, consider column (e) in Fig. 3.10 as an example. The rigid regions including nose, eyes and forehead altogether possess about 56% importance of the whole face, while the mouth region has only about 2.8% importance. It is worth noting that the facial cheek regions (on both sides), which are usually considered as non-rigid regions, own more than 20% importance, showing that much identity related information also exists in those non-rigid facial regions. Note that these kinds of weight distribution patterns are quite different from those of the 2D face, especially in the nose region in comparison with Fig. 3.1 (b). This difference can be explained by the different nature of the data between 2D and 3D faces; for example, the nose region in a 2D image is easily influenced by the variations in lighting whilst that of the 3D face remains stable during variations in expression.

To evaluate the effectiveness of W-SRC in facial expression variations, we compare the performance of SRC and W-SRC on FRGC v2.0, Bosphorus, and BU-3DFE respectively (see Table 3.6, 3.7, and 3.8). The weights learned from FRGC v1.0, BU-3DFE, and Bosphorus are denoted as F-W-SRC, BU-W-SRC, and BO-W-SRC respectively. The local normal encoding operator $Q_{2,16}$ is used in all three Tables. Table 3.6 presents the rank-one scores on the FRGC v2.0 database using SRC, F-W-SRC, BU-W-SRC, and BO-W-SRC. The results using the single normal component LNPx($Q_{2,16}$), LNPy($Q_{2,16}$), and LNPz($Q_{2,16}$) and that of their fusion MC-LNP($Q_{2,16}$) are reported. We can see from Table 3.6 that the performance of F-W-SRC is slightly better than SRC except LNPx($Q_{2,16}$). The results of BU-W-SRC and BO-W-SRC are similar and both are improved by 1.5% to

Table 3.6: *Comparison of rank-one score improvements on the FRGC v 2.0 database: patch weights are learned using FRGC v1.0, BU-3DFE, and Bosphorus respectively.*

	LNPx($Q_{2,16}$)	LNPy($Q_{2,16}$)	LNPz($Q_{2,16}$)	MC-LNP($Q_{2,16}$)
SRC	87.01%	86.13%	88.43%	92.60%
F-W-SRC	86.63%	88.40%	88.65%	93.59%
BU-W-SRC	**88.85%**	88.54%	**90.58%**	94.50%
BO-W-SRC	88.62%	**88.88%**	90.41%	**94.61%**

Table 3.7: *Comparison of rank-one score improvements on the BU-3DFE database: patch weights are learned from Bosphorus.*

	LNPx($Q_{2,16}$)	LNPy($Q_{2,16}$)	LNPz($Q_{2,16}$)	MC-LNP($Q_{2,16}$)
SRC	**78.92%**	80.92%	84.08%	88.25%
BO-W-SRC	78.83%	**84.29%**	**86.21%**	**90.71%**

2% in comparison with SRC. These results suggest W-SRC used with the weight learning strategy does provide more robustness to facial expression variations than SRC.

Table 3.7 presents the rank-one scores on the BU-3DFE database using SRC and BO-W-SRC. We can see that the performance improvements based on BO-W-SRC are largely different in the three normal components, with -0.09%, 3.37% and 2.13% for LNPx, LNPy and LNPz respectively. These results indicate that the facial surface deformations caused by facial expression variations are likely to break down into different quantities in different normal components. The improvement of the fusion result using MC-LNP is about 2.5% which also proves the effectiveness of W-SRC handling facial expression variations.

Table 3.8 presents the rank-one scores on the Bosphorus database using SRC and BU-W-SRC. We can see that BU-W-SRC improves the performance for all three normal components, with 3.76%, 2.03% and 4.19% for LNPx, LNPy, and LNPz respectively.

Table 3.8: *Comparison of rank-one score improvements on the Bosphorus database: patch weights are learned from BU-3DFE.*

	LNPx($Q_{2,16}$)	LNPy($Q_{2,16}$)	LNPz($Q_{2,16}$)	MC-LNP($Q_{2,16}$)
SRC	83.12%	86.24%	84.91%	90.92%
BU-W-SRC	**86.88%**	**88.27%**	**89.10%**	**93.21%**

3.5.3.5 Experiment V: Comparison of the performance degradation due to facial expression variations

We first evaluate the degradation influenced by facial expression variations of the FRGC v2.0 database. According to the experimental protocol used in [Mian *et al.* 2008], [Huang *et al.* 2010] and [Huang *et al.* 2011a], we split all probe faces into two subsets based on their original expression labels. The first subset consists of only neutral faces, while the second one is only made up of non-neutral faces. The performance degradation, reflected by the difference between the accuracies of subset I and II reported in Table 3.9, is utilized to analyze the robustness to facial expression variations. We can see from this table that the 6.6% drop is achieved based on the proposed MSMC-LNP descriptor and SRC, and the 3.8% drop is obtained by using Bosphorus database as the training set for W-SRC. Note that our performance on subset I is a little worse than [Mian *et al.* 2008, Huang *et al.* 2010, Huang *et al.* 2011b], while the degradations are competitive in comparison with their results.

Table 3.9: *Comparing the degradations of rank-one scores influenced by facial expression changes on the FRGC v 2.0 database (Subset I: neutral probes; Subset II: non-neutral probes).*

	Sub. I	**Sub. II**	**Degradation**
(1) Mian et al. [Mian *et al.* 2008]	99.0%	86.7%	12.3%
(2) Huang et al. [Huang *et al.* 2010]	99.1%	92.5%	6.6%
(3) Huang et al. [Huang *et al.* 2011a]	99.0%	94.9%	4.1%
(4) MSMC-LNP + SRC	97.1%	90.5%	6.6%
(5) MSMC-LNP + BO-W-SRC	98.0%	94.2%	**3.8%**

Since all the 2,400 non-neutral probe faces in the BU-3DFE database have labels of expression intensity levels (increasing from level 1 to level 4), it is thus possible to quantify the performance drop as the intensity of facial expressions increases. For this purpose, all the probe faces are divided into four subsets according to their labels of expression intensity. Subset I, II, III, and IV are made up of the probe faces with the expression intensity of level 1, level 2, level 3, and level 4 respectively, and each subset consists of 600 probe faces with six prototypical expressions.

Table 3.10: *The degradations of rank-one scores of the proposed approach as the intensity of facial expressions increases in the BU-3DFE database.*

	Sub. I	**Sub. II**	**Sub. III**	**Sub. IV**
MSMC-LNP + SRC	97.0%	94.0%	90.5%	80.5%
MSMC-LNP + BO-W-SRC	**97.3%**	**95.0%**	**92.7%**	**83.8%**

The performance is shown in Table 3.10. We can find that the degradation from the lower level to higher level expression intensity becomes larger and larger especially from Subset III to Subset IV. By using SRC without weight learning, the degradations are 3.0%

from Subset I to Subset II, 3.5% from Subset II to Subset III, and 10.0% from Subset III to Subset IV. By using BO-W-SRC, the degradations are 2.3% from Subset I to Subset II, 2.3% from Subset II to Subset III, and 8.9% from Subset III to Subset IV. As we can see, all three degradations using BO-W-SRC are smaller than those using SRC, suggesting that the weight learning of different local facial regions is an effective way to improved robustness to expression changes.

3.5.3.6 Experiment VI:The performance of distinguishing between identical twins across expression variations

In this experiment, we evaluate the performance of our system to distinguish between identical twins with a smiling expression. We regard the SRC based recognition rate as the baseline and compare it with W-SRC, where the patch weights are learned from different training sets. Given the fact that there are only neutral and smile scans in the 3D-TEC dataset, a specific training set is designed based on the subset of Bosphrous dataset, *i.e.* 105 first neutral scans as gallery and 105 happy scans as probe, the corresponding W-SRC is denoted as BOS-W-SRC. All the rank-one scores achieved by using MSMC-LNP feature as well as SRC, F-W-SRC, BU-W-SRC, BO-W-SRC, and BOS-W-SRC classifiers are shown in Table 3.11.

Table 3.11: *Comparison of the rank-one scores on 3D-TEC by using different training sets.*

Algorithm	Rank-one scores			
	I	II	III	IV
MSMC-LNP + SRC	**94.9%**	96.3%	89.3%	88.3%
MSMC-LNP + F-W-SRC	93.5%	94.4%	88.8%	88.3%
MSMC-LNP + BU-W-SRC	93.9%	96.3%	90.7%	91.6%
MSMC-LNP + BO-W-SRC	94.4%	**96.7%**	**90.7%**	**92.5%**
MSMC-LNP + BOS-W-SRC	**95.8%**	**96.7%**	**95.3%**	**95.3%**

From Table 3.11, we can see that the performance improvements are very limited for F-W-SRC, BU-W-SRC, and BO-W-SRC. The main reason is the asymmetry of the training and testing data. The 3D-TEC dataset contains identical twin samples with only neutral and smiling expressions, while the FRGC v1.0 database only includes neutral expression scans; BU-3DFE and Bosphrous databases consist of the scans with different expression types. The performance of BOS-W-SRC confirms this analysis, *i.e.* when the sample distributions of the training and testing sets are more similar to each other, W-SRC will be more efficient, with 6% and 7% improvements for Cases III and IV.

3.5.3.7 Experiment VII: Comparison with the state-of-the-art

To further evaluate the performance of the proposed method, in Table 3.12 we display a comprehensive comparison of the rank-one recognition rates of the FRGC v2.0, Bosphrous, BU-3DFE, and 3D-TEC databases with the state of the art. In this table, our best results

Table 3.12: *Comparison of the Rank-1 scores on the FRGC v2.0, Bosphrous, BU-3DFE and 3D-TEC databases.*

Approaches	FRGC v2.0	Bosphrous	BU-3DFE	3D TEC			
				Case I	Case II	Case III	Case IV
(1) Chang *et al.* [Bowyer *et al.* 2005]	91.9%	-	-	-	-	-	-
(2) Cook *et al.* [Cook *et al.* 2006]	92.9%	-	-	-	-	-	-
(3-a) Mian *et al.* [Mian *et al.* 2008]	93.5%	-	-	-	-	-	-
(3-b) Mian *et al.* [Mian *et al.* 2007]	96.2%	-	-	-	-	-	-
(3-c) Mian *et al.* [Al-Osaimi *et al.* 2008]	93.8%	-	-	-	-	-	-
(3-d) Mian *et al.* [Al-Osaimi *et al.* 2009]	**96.5%**	-	-	-	-	-	-
(4) Abate *et al.* (Normal) [Abate *et al.* 2009] (1024 faces)	92.2%	-	-	-	-	-	-
(5) Tang *et al.* (Normal + Depth) [Tang *et al.* 2013] (900 faces)	94.89%	-	-	-	-	-	-
(6-a) Faltemier *et al.* (ICP) [Faltemier *et al.* 2008a]	**97.2%**	-	-	-	-	-	-
(6-b) Faltemier *et al.* (ICP) [a] [Faltemier *et al.* 2008a]	**98.0%**	-	-	93.5%	93.0%	72.0%	72.4%
(6-c) Faltemier *et al.* (ICP) [b] [Faltemier *et al.* 2008a]	**98.0%**	-	-	94.4%	93.5%	72.4%	72.9%
(7-a) Huang *et al.* (SI) [Huang *et al.* 2010]	91.8%	-	-	92.1%	93.0%	83.2%	83.2%
(7-b) Huang *et al.* (eLBP) [Huang *et al.* 2011a]	**97.2%**	-	-	91.1%	93.5%	77.1%	78.5%
(7-c) Huang *et al.* (Range PFI) [Huang *et al.* 2011d]	95.5%	-	-	91.6%	93.9%	68.7%	71.0%
(8) Gokbert *et al.* (Normal + Geometry) [Gokberk *et al.* 2006]	87.8%	-	-	62.6%	63.6%	54.2%	59.4%
(9-a) Kakadiaris *et al.* (Normal + Geometry) [Kakadiaris *et al.* 2007a] [Ocegueda *et al.* 2011b]	**97.0%**	-	-	**98.1%**	**98.1%**	91.6%	93.5%
(9-b) Kakadiaris *et al.* (Normal) [c] [Ocegueda *et al.* 2011a]	**97.9%**	**98.2%** (2797/105)	**99.7%**	-	-	-	-
(10) Alyuz *et al.* [Alyuz *et al.* 2010]	**97.5%** (94.55%)	**98.2%** (2814/105)	-	-	-	-	-
(11) C. Maes *et al.* [Smeets *et al.* 2013]	89.6%	**97.7%** (3186/105)	-	-	-	-	-
(12) H. Li *et al.* [**?**]		94.1% (4561/105)	-	-	-	-	-
(13) Queirolo *et al.* [Queirolo *et al.* 2010]	**98.4%**	-	-	-	-	-	-
(14) Spreeuwers *et al.* [Spreeuwers 2011]	**99.0%**	-	-	-	-	-	-
MSMC-LNP + SRC	94.1%	-	-	94.9%	96.3%	89.3%	88.3%
MSMC-LNP + BU-W-SRC	**96.3%**	**95.4%** (2797/105)	-	93.9%	96.3%	90.7%	91.6%
MSMC-LNP + BO-W-SRC	**96.3%**	-	**92.21%**	94.4%	96.7%	90.7%	92.5%
MSMC-LNP + BOS-W-SRC	-	-	-	**95.8%**	**96.7%**	**95.3%**	**95.3%**

[a] match scores normalization using E_{pkn}, $E_{pkn}(p,g_k) = E_min(p,g_k)/\sum_{j=1,j\neq k}^{N}(E_min(g_i,g_k)/(N-1))$, where p is a probe image, g_k are the gallery images, and N is the number of gallery images. $E_min(p_1,p_2) = min(E(p_1,p_2),E(p_2,p_1))$, and $E(p_1,p_2)$ be the match score of point clouds p_1 and p_2.

[b] match scores normalization using E_{minmax}, which is the min-max normalization on the resulting match score from the E_{pkn} normalization

[c] Bosphrous, BU-3DFE and Bosphrous as the training sets corresponding to FRGC v2.0, Bosphrous and BU-3DFE as the testing sets respectively.

are highlighted as well as those of the state of the art which are better than ours. From this table, we can observe that:

(i) The vast majority of 3D FR algorithms only report their performance on FRGC v2.0 which is the de facto standard benchmark in 3D FR. Table 3.12 lists only some of them. However, because of different nature of 3D face data captured in different conditions and scenarios, a 3D FR technique with high performance on FRGC v2.0 does not guarantee the same level of performance on other 3D face data and can even vastly degrade, *e.g.*, methods (6-b), (6-c), (7-a), (7-b), (7-c) and (8) on 3D TEC. Note that for method (10), the LDA approach has a poor generalization ability. The rank-one score 97.5% achieved on the FRGCv2.0 database will drop to 94.55% when the Bosphorus database instead of FRGCv1.0 is used to constructed the LDA space. This means that it is important to benchmark 3D FR techniques on other datasets to check their generalizability. To the best of our knowledge, except for our method, only (9-a) and (9-b) report their results on all the four databases.

(ii) The proposed approach, *i.e.* Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP) along with the weighted SRC for an improved robustness to facial expression variations, consistently displays competitive rank-one recognition rates on these various datasets, and even the best results in the difficult scenarios, namely case III and case IV with 3D-TEC, and thereby demonstrates its effectiveness and its generalizability, in particular with respect to various types and intensities of facial expressions. The results achieved by the proposed approach on 3D-TEC tend to suggest that the proposed MSMC-LNP facial feature is discriminative enough to distinguish between even identical twins despite their strong similarities whereas W-SRC, in learning the discriminative weights of different facial regions, makes the proposed solution robust enough to facial expression variations as stated in the difficult cases III and IV on 3D-TEC.

(iii) Facial normal information is also used as 3D facial representation for the extraction of facial features, *e.g.*, methods (4), (5), (8), (9-a) and (9-b). In (4), the difference of normal maps is used as similarity measurement, while a rank-one score of 92.2% is reported on a subset of FRGC v2.0 (1024 samples) database; In (5), both facial depth and normal information are encoded using LBP. As regards the encoding of normal information, they propose an original method which encodes the angle difference between the normal vector of a given central vertex and those of its eight neighboring ones over facial triangular surfaces. They further make use of 20 manually labelled facial markers to build a 2D specific feature-based sparse division and reported a rank-one score of 94.89% on a subset of FRGC v2.0 (900 samples of 150 subjects) under a different experimental evaluation protocol. In (8), Gokbert *et al.* use surface normal variance at each pixel location as a distance measure between face images and report a rank-one score of 87.8% on the whole FRGC v2.0 database, while this reasonable performance vastly deteriorates on the 3D-TEC database, achieving only around 60% rank-one recognition rate. In (9-a) and (9-b), wavelet coefficients are used as a similarity measurement on both normal and geometry maps. Note that this method is the one of the best 3D FR methods in the literature with a very good generalization ability. Kakadiaris *et al.* in (9-a) and (9-b) use a very sophisticated multistage face registration (spin images, ICP, and Simulated Annealing on Z-Buffers) and deformable model fitting techniques, and a Linear Discriminant Analysis (LDA) based feature selection technique in their following works [Ocegueda *et al.* 2011b, Ocegueda *et al.* 2011a]. Compared with the proposed method, they make use of additional information of 3D face scans,

e.g., geometry image, and achieve better results on all the four databases except difficult Case III and Case IV over the 3D-TEC database.

In addition to recognition accuracy and good generalizability, processing time is another key factor in judging the effectiveness of a 3D FR algorithm. Table 3.13 lists the computation time of our method and a number of top-ranking 3D FR methods on the FRGC v2.0 database. For each method, we report the time required for preprocessing and/or feature extraction, face matching, and identification of a single probe using a gallery of 466 subjects of the FRGC v2.0 database in the first three columns. In the last column, we also report the rank-one recognition rates of different approaches. For our method, all range images are resized to 120× 96 and it takes 2.4 seconds for registration, 0.65 seconds for the extraction of a LNP-based facial feature, whereas normal estimation takes 0.62 seconds and normal encoding takes 0.03 seconds for each scale and each normal component. For face matching, unlike other matching methods which compare the probe face with every gallery face, the SRC-based method compares a probe face with all the gallery faces at the same time by solving a l_0 minimization problem. The computation complexity of this optimization problem is related to the dictionary size (*i.e.*, gallery size) and the feature dimensionality. In our case, it takes 0.5 seconds to match a probe face and all the 466 galley faces with the highest encoding dimensionality. Thus, our method only takes 3.55 seconds to identify a single probe in a gallery of 466 subjects such as those in the FRGC v2.0 database. This makes the proposed method much faster than the competing methods except [Spreeuwers 2011]. However, the generalizability of the 3D FR method in [Spreeuwers 2011] is not checked on other datatsets as we did. Note that our method is implemented in Matlab and the computation time of our method tested on a PC with Intel Core 2 CPU and 2.66 GHz. Please refer to [Spreeuwers 2011] for more information regarding the processor and the computation time of all the other methods in the table.

Table 3.13: *Computation time for preprocessing and/or feature extraction, face matching and identification of a single probe using a gallery of 466 subjects of the FRGCv2 database.*

Approach	Preprocessing (s)	Matching (s)	Total time (s)	Rank-one
Spreeuwers *et al.* [Spreeuwers 2011]	2.5	1/ 11 150	**2.5**	99.0%
Queirolo *et al.* [Queirolo *et al.* 2010]	-	4	1864	98.4%
Alyuz *et al.* [Alyuz *et al.* 2010]	131	1/ 20 000	131	97.5%
Faltemier *et al.* [Faltemier *et al.* 2008a]	7.5	2.4	1312	97.2%
Kakadiaris *et al.* [Kakadiaris *et al.* 2007a]	15	1/1000	15.5	97.0%
Mian *et al.* [Al-Osaimi *et al.* 2009]	4	10	50.6	96.5%
This work	3.05	0.5	**3.55**	96.3%

3.6 Conclusion and Future Work

In this chapter, we have presented an expression-robust 3D face identification approach based on a novel 3D facial surface descriptor, namely Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP), along with a weighted Sparse Representation-based Classifier (W-SRC). The extensive experiments that we carried out using 4 different 3D face datasets indicate that: 1) LNP is much more discriminative than either the original

normal information or LSP. 2) SRC is more efficient than the Chi-square distance based classifier. 3) Both multi-scale and multi-component are efficient ways of improving the performance of LNP. 4) The importance of local facial regions for 3D face identification is quite different from the 2D-based one, especially in the nose region. 5) Patch-weight based W-SRC is very robust to facial expression variations, even for identical twins with expression changes, and a significant improvement can be achieved if the distributions of training and testing sets are similar to each other. 6) The proposed system (i.e. MSMC-LNP + W-SRC) consistently achieves competitive rank-one recognition rates on the FRGC v2.0, Bosphrous, BU-3DFE, and 3D-TEC databases and thereby displays a good generalizability with respect to various types of facial expressions. 7) In addition to competitive recognition accuracy and good generalizability, the proposed method is also very fast.

In the future, we will focus on the following three aspects to further improve the proposed method. 1) In this work, a basic 3D face alignment method, *i.e.* few iterations of ICP with respect to a reference model, was used. Recent works [Spreeuwers 2011, Mohammadzade & Hatzinakos 2013] suggest that a better 3D face alignment can further lead to significant performance improvement. We are seeking to improve the proposed approach with a better alignment method; 2) In this work, we presented only the identification results of the proposed method. Notice that there are very few works on the study of sparse representation-based classifiers for 2D and 3D face verification. Reference [Guo *et al.* 2012] is perhaps the only work stressing this issue for 2D face verification. In our future work, we will explore the sparse representation-based classifiers for 3D face verification and report the verification results on the FRGC v2.0, Bosphrous, BU-3DFE and 3D-TEC databases. 3) In this work, only the challenge of facial expression variations was considered. However, in real-life applications, 3D face scans concern uncooperative subjects in uncontrolled conditions and in arbitrary poses, thus with missing face data, along with possible external occlusions (*e.g.*, glasses, scarfs, *etc.*). We want to extend this work to face these new challenges.

CHAPTER 4

3D Face Recognition in the Presence of Expression, Pose and Occlusion Variations

Contents

4.1 Introduction . . . 98
4.1.1 Overview of the Proposed Method . . . 100
4.1.2 Main Contributions of the Proposed Method . . . 101
4.2 Related Work . . . 103
4.2.1 3D Facial Representations . . . 103
4.2.2 Facial Features . . . 104
4.2.3 Matching Scheme . . . 105
4.3 Estimating Multiple Order Facial Surface Differential Quantities . . . 106
4.3.1 Facial Surface Gradient Estimator . . . 106
4.3.2 Facial Surface Curvature Estimator . . . 107
4.3.3 Facial Curvature Derivation Estimator . . . 108
4.4 3D Facial Salient Vertices Detection and their Local Shape Description 109
4.4.1 3D Facial Salient Vertices Detection: κ_{max} and κ_{min} based Detectors 109
4.4.2 Pose-invariant Local Shape Description: Histograms of Multi-order Surface Differential Quantities . . . 110
4.5 Facial Surface Matching: a SRC-based Fine Grained Matcher vs. a Coarse Grained Matcher . . . 112
4.5.1 The Coarse Grained Matcher (CGM) . . . 113
4.5.2 SRC-based Fine Grained Matcher (FGM) . . . 113
4.6 Experimental Results . . . 116
4.6.1 Databases . . . 116
4.6.2 Experimental Setups . . . 119
4.6.3 Analysis of Salient Vertices Located by κ_{max} and κ_{min} Detectors . . . 121
4.6.4 Face Recognition on the Bosphorus Database . . . 124
4.6.5 Face Recognition on the FRGC v2.0 Database . . . 133
4.6.6 Discussion . . . 136
4.7 Conclusion . . . 141

4.1 Introduction

Face is potentially one of the best biometrics for people identification related applications, since it is non-intrusive, contactless and socially well accepted. The past three decades have witnessed tremendous research effort firstly focused on 2D face images [Zhao *et al.* 2003] and more recently on 3D face scans [Bowyer *et al.* 2006]. Despite the great progress achieved so far in the field [Zhao *et al.* 2003], face recognition (FR) in 2D are still not reliable enough [Abate *et al.* 2007b] in uncontrolled environment, especially in the presence of pose and lighting changes [Phillips *et al.* 2000]. Along with the rapid development of 3D imaging systems, there has been a lot of research interest in 3D face analysis and recognition over the past decade for an improved robustness to pose and lighting variations in comparison with 2D FR. However, the overwhelming majority of research works so far in the literature has targeted 3D FR using facial expressive but frontal 3D face scans (*e.g.*, FRGC datasets), although 3D face scans captured in a unconstrained environment can depict not only facial expression variations but also arbitrary pose changes which can lead to self-occlusion and thereby large missing face data. Moreover, they can also be severely altered by external occlusions, *e.g.*,hand, scarf, hair, glasses, *etc.* Fig. 4.1 illustrates these diverse challenges using several 3D face scans of the *Bosphorus* dataset [Savran *et al.* 2008].

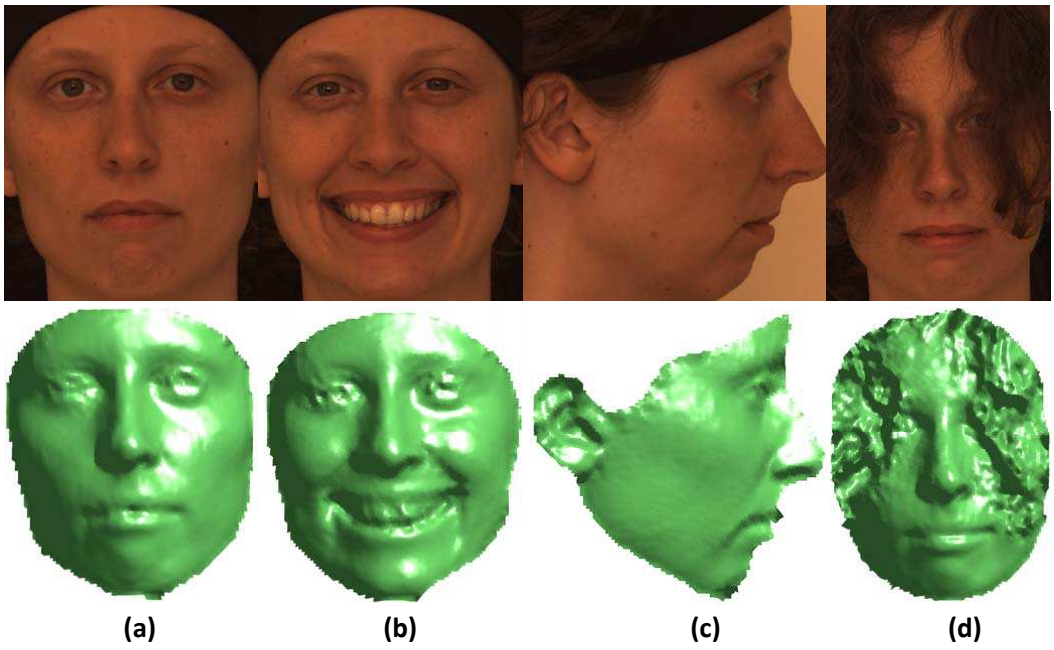


Figure 4.1: *2D texture images (1^{st} row) and 3D face scans (2^{nd} row) from the Bosphorus dataset illustrating the various challenges in real-life applications under uncontrolled environment: (a) frontal and neutral; (b) facial expression variation; (c) pose variation and data missing due to self-occlusion; and (d) external occlusion.*

Depending on the underlying application requirements, 3D FR techniques can be designed differently based on the underlying applicative scenario, *i.e.* identification or verifi-

cation, and the assumption whether users are cooperative or not under controlled environment or not. In the scenario of verification, a subject claims an identity to be checked by a biometric system, the subject is generally assumed cooperative so that the claimed identity be accepted by the system. The scenario of identification requires identifying a probe face among the gallery ones and can take place either in controlled conditions with cooperative users or uncontrolled conditions with uncooperative users. Although one can envisage low quality 3D face scans (*e.g.,* rotated and occluded faces) in the gallery, the overwhelming majority of works in the literature assumes a neutral frontal 3D face scan of good quality per subject in the gallery. When users are assumed cooperative, a frontal and neutral face scan of good quality, thus with moderate pose variations, can be acquired for probe. The major challenge in this case is therefore to design an appropriate matching scheme which is robust enough to tolerate possible intra-class facial shape variations over time (*e.g.*, face morphology changes with age) and fine enough to distinguish extra-class 3D face similarities (*e.g.*, identical twins). However, when users are assumed uncooperative in uncontrolled settings, thereby within a scenario of identification, a probe 3D face scan can display any facial expression, thus leading to significant facial surface deformation, be in an arbitrary pose, thereby possibly resulting in large missing face data, and occluded by external objects (*e.g.*, sun glasses, scarf, hands, *etc.*). In such a case, 3D FR algorithms have to face these multiple challenges within the same framework.

However, while 3D face data is widely expected to help overcome the unsolved issues in 2D FR, *i.e.* lighting variations and pose changes, which typically occur in uncontrolled conditions, the overwhelming majority of 3D FR algorithms so far proposed in the literature only target the challenge of facial expressions using frontal face data with slight pose variations as those in the FRGC v2.0 dataset [Phillips *et al.* 2005]. Very high performance, up to 99.0% rank-1 recognition rate, has been reported on FRGC v2.0 by state of the art 3D FR algorithms which mostly operate on range face data, thereby require a more or less accurate and costly face registration step (*e.g.*, ICP), through several anthropometric landmarks (*e.g.*, nose tip), to align frontal 3D scans with moderate pose variations [Kakadiaris *et al.* 2007a] [Faltemier *et al.* 2008a] [Queirolo *et al.* 2010]. Those methods reporting a performance among the best on FRGC v2.0 may also require an additional stage of learning to further boost the performance [Wang *et al.* 2010] [Spreeuwers 2011]. Using the benchmark of FRGC v2.0, Kakadiaris *et al.* [Kakadiaris *et al.* 2007a] reported a high performance at the early stage of its public release with a method which performs a multi-stage fine alignment to fit a deformable Annotated Face Model (AFM) to an input face scan. They further improved their approach using a LDA-based learning strategy to select highly discriminating wavelet coefficients extracted from the geometry image which results from the fitting of AFM [Ocegueda *et al.* 2011a]. Queirolo *et al.* [Queirolo *et al.* 2010] reported a 98.4% rank-1 recognition rate with a method which first locates on a frontal range image six anthropometric landmarks, *e.g.*,the inner right and left eye corners, the right and left nose corners, and the nose tip and base, and then crop four face regions on which a boosted ICP using Simulated Annealing is performed for matching.

Spreeuwers [Spreeuwers 2011] reported a rank-1 recognition rate as high as 99.0% with a method which aligns 3D face scans through an intrinsic face coordinate system thanks to landmark structures, *i.e.* the vertical symmetry plane, the nose tip and the slope of the bridge of the nose, then fuse several region classifiers trained on additional 3D face data, *e.g.*, Bosphorus, using subspace-based methods. Unfortunately, all these methods

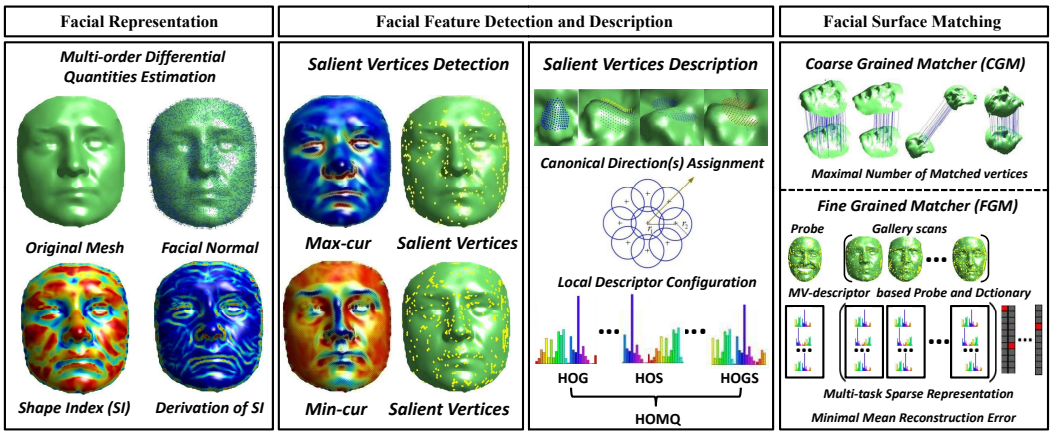


Figure 4.2: *Overview of the proposed method.*

implicitly assume that 3D face scans are captured at a frontal pose and thereby make it possible the use of a priori face information, *e.g.*, nose tip, symmetry plane, that may be simply unavailable because of face missing data due to pose or external occlusions in a scenario of 3D FR under uncontrolled conditions.

There does exist in the literature a few methods dealing with one specific or another challenge of 3D FR in uncontrolled conditions, *e.g.*, pose estimation using the nose tip direction [Dibeklioglu *et al.* 2009] or facial landmark structure [Passalis *et al.* 2011], the recovery of missing face data due to pose using facial symmetry plane [Passalis *et al.* 2011], the detection and restoration of occluded facial regions in aligning an input potentially occluded 3D face scan with some landmarks and the learning of a subspace of registered frontal 3D face scans [Colombo *et al.* 2011a] [Drira *et al.* 2013]. However, it is hard to predict in advance the specific kind of degradation on an input 3D face scan captured from a uncooperative subject in uncontrolled conditions. The possible joint presence of pose changes and external occlusions makes it difficult to automatically localize facial landmarks or structures which are required by the previous techniques for face registration.

In this chapter, we are concerned with 3D FR under uncontrolled conditions and therefore target an identification scenario dealing with those aforementioned challenges. For this purpose, we propose a registration free local feature-based 3D FR method which performs the matching of two 3D face scans through their respective shape description, *i.e.* a sparse set of pose invariant local shape descriptors, and thereby offers a uniform framework to deal with the challenges of facial expression variations, pose changes and external occlusions at same time.

4.1.1 Overview of the Proposed Method

As shown in Fig. G.1, given an input probe 3D face scan, we first estimate its multiple order surface differential quantities, including facial normal (1^{st} order), facial shape index (2^{nd} order) and their derivation (3^{rd} order). Then, facial salient vertices are detected by the local extreme values of maximum curvature (κ_{max}) and minimum curvature (κ_{min}) within a 3D Gaussian scale space, giving birth to κ_{max} salient vertex detector and κ_{min} salient vertex detector. Once salient vertices located on the input probe scan, we move

to the description of local shapes around those salient vertices and compute histograms of multiple order surface differential quantities, including Histograms of mesh Gradient (HOG, 1^{st} order), Histogram of Shape index (HOS, 2^{nd} order) and Histogram of Gradient of Shape index (HOGS, 3^{rd} order), and their early fusion, namely Histogram of Multiple Quantities (HOMQ, multi-order). The input probe face scan is then represented as a sparse set of pose invariant local shape descriptors computed over those automatically located salient vertices, leading to Multiple Vertices based HOG, HOS, HOGS and HOMQ, denoted as MV-HOG (1^{st} order), MV-HOS (2^{nd} order), MV-HOGS (3^{rd} order) and MV-HOMQ (multi-order), respectively. The same pipeline is also applied once to all the 3D face scans in the gallery. The matching of a probe 3D face scan and a gallery one is carried out through their respective sparse set of local shape descriptors, using a SRC-based fine grained matcher (FGM) which computes the mean reconstruction error. FGM is further compared with a coarse grained matcher (CGM) which simply counts the number of matched pairs of salient vertices between the probe and gallery.

4.1.2 Main Contributions of the Proposed Method

1) Three novel pose-invariant local shape descriptors are proposed which describe the neighborhood of a given salient vertex as weighted statistical distributions of multiple order surface differential quantities, including Histograms of mesh Gradient (HOG, 1^{st} order), Histogram of Shape index (HOS, 2^{nd} order) and Histogram of Gradient of Shape index (HOGS, 3^{rd} order). As each of those local shape descriptors provides complementary geometric shape information, we also consider a local shape descriptor, namely Histogram of Multiple Quantities of multiple order (HOMQ), which is their early fusion, *i.e.* their simple concatenation;

2) Automatic extraction of a sparse set of salient vertices which are detected as the local extreme values of maximum and minimum curvatures within a 3D Gaussian scale space. These salient vertices prove to be evenly distributed on both the rigid and elastic facial regions, thereby achieve a comprehensive but still compact description of a facial surface when combined with the proposed local shape descriptors. The fact that these salient vertices cover the whole face surface is important because we believe that both rigid and elastic facial regions carry out discriminating information. The facial description which results from such a process remains compact as the detected salient vertices are quite sparse and their number typically around 650 in comparison with the number of vertices of a raw point cloud which is on the order of multiple ten thousands, *e.g.*, 30,000. Each facial surface is thus represented as a sparse set of local shape descriptors, namely Multiple Vertices based HOG, HOS, HOGS and HOMQ, denoted as MV-HOG (1^{st} order), MV-HOS (2^{nd} order), MV-HOGS (3^{rd} order) and MV-HOMQ (multi-order), respectively. Such a compact facial description prevents one to recover the original raw 3D face data which is in line with the respect of privacy policy as required in practical biometric applications;

3) A fine grained local feature matcher (FGM) using the Sparse Representation based Classifier (SRC) [Wright *et al.* 2009a] is introduced and compared with a coarse grained local feature matcher (CGM) that was used in our preliminary work [Li *et al.* 2011c]. CGM performs the matching of two face scans through their corresponding sparse sets of local shape descriptors and simply counts the number of matched vertex pairs as matching score. This is problematic when the underlying face scans have severe missing data due to pose and/or external occlusions, which result in a significant decrease of the number of matched

vertex pairs. In contrast, FGM alleviates this vulnerability of missing data by making use of SRC to compute a mean reconstruction error when each local shape descriptor from a probe face scan is sparsely reconstructed from an over-complete dictionary composed of the local shape descriptors extracted from all the face scans in the gallery.

4) All the previous steps are directly carried out on mesh-based 3D face scans without any prior registration or learning. The overall result is a uniform registration free 3D FR framework which can deal with facial expression variations, arbitrary pose changes and occlusions at the same time. This is in clear contrast to the overwhelming majority of 3D FR algorithms which only deal with expressive frontal face scans and mostly require in addition an expensive registration step.

Comprehensive experiments were carried out on Bosphorus, the largest public dataset which best accounts for the challenges of 3D FR in uncontrolled conditions [Savran *et al.* 2008] and depicts large variations of expressions, pose and occlusions [Savran *et al.* 2008]. Indeed, Bosphorus encompasses 4666 3D face scans of 105 subjects depicting 34 expressions (both action units and six basic emotions), 13 pose variations (yaw, pitch and cross rotations) leading in particular to self-occlusion and thus large missing data, and 4 external occlusions (*e.g.*,hand, hair, eyeglasses). It permits thus to comprehensively test the robustness of 3D FR algorithms against the aforementioned challenges in uncontrolled conditions. Without any prior registration or learning, the experimental results demonstrate the effectiveness of the proposed method which achieves the best state of the art rank-1 recognition rates over the whole dataset (96.56%) and its three subsets of expression variations (98.82%), pose changes (91.14%) and occlusions (99.21%). As recent studies show that state of the art 3D FR algorithms with high performance over a database can vastly degrade over 3D face scans captured under different conditions in a different scenario [Vijayan *et al.* 2011b], we also carried out additional experiments on the *FRGC v2.0* database. The experimental results show that the proposed method consistently displays competitive performance in comparison with state of the art algorithms which mostly require accurate face alignment and/or learning of discriminating features, and thereby demonstrates its generalization skill.

In our preliminary work [Li *et al.* 2011c], we assumed that both rigid and elastic facial regions have contributions to 3D face recognition, and thus proposed a local matching approach to robustly associate the descriptors extracted from salient vertices which are detected on the entire facial surface in the presence of expression variations. With it we took part in the track on 3D face retrieval and recognition at SHREC 2011 contest, and our two runs were awarded the first and the second performance in terms of rank one recognition rate out of 14 runs submitted by four research groups [Veltkamp *et al.* 2011]. Furthermore, since the probe set contains a certain number of partial face scans, our method also presents the potential to deal with data missing caused by severe pose changes as often occurs in the real. This work integrates that preliminary work but significantly extends it. The main extensions are summarized as follows: (i) Instead of a coarse grained matching scheme as defined in [Li *et al.* 2011c] which simply counts the number of matching vertices in a pair of 3D face scans, we introduce in this paper a fine grained matching scheme based on a multi-task sparse representation to further boost the robustness of the proposed approach with the challenges of real-life applications in uncontrolled settings; (ii) In addition to the challenge of facial expression variations, we further consider the two other challenges, namely pose changes and occlusions through the improved framework and benchmarked

the proposed framework on the whole Bosphorus dataset and its various subsets. Moreover, we also provide in-depth analysis of the proposed method compared with the state of the art; (iii) Additional experiments were also carried out on the *FRGC v2.0* database to check the generalization skill of the proposed method in comparison with the state of the art. (iv) We also give insights into and provide in-depth analysis of the proposed method as well as the time complexity analysis.

The remainder of this chapter is organized as follows: Section 2 overviews the related work. Section 3 introduces multiple order surface differential quantities which are estimated on facial triangular meshes. Section 4 presents the detection of salient vertices and the construction of various pose invariant local shape descriptors. Section 5 describes the SRC-based fine grained matcher (FGM) along with the coarse grained matcher (CGM). Experimental results are discussed in section 6. Section 7 concludes this chapter.

4.2 Related Work

3D FR literature has been extensively surveyed by Bowyer *et al.* [Bowyer *et al.* 2006] for works appeared prior to 2006, and Huang *et al.* [Huang *et al.* 2012], Smeets *et al.* [Smeets *et al.* 2012], Spreeuwers [Spreeuwers 2011] and Drira *et al.* [Drira *et al.* 2013] for more recent contributions. In this section, we try to broadly categorize them and analyze their advantages and shortcomings, in particular with respect to the challenges of 3D FR in uncontrolled environments, *i.e.* facial expressions, pose and occlusions, in order to introduce the proposed approach.

A typical 3D FR algorithm comprises the following major components although they are strongly interwoven each other [Gokberk *et al.* 2008]: selection of a proper face representation, feature extraction, the design of a matching scheme which can be possibly followed by a fusion strategy. As a result, 3D FR algorithms can be categorized using a number of criteria, depending on the underlying technique chosen for each of these components.

4.2.1 3D Facial Representations

3D FR algorithms can be first differentiated by the chosen face representation from which facial features can be extracted for the matching step. Different face representations were utilized in the literature, including raw point cloud [Faltemier *et al.* 2008a], range image [Spreeuwers 2011], or more elaborated representations, *e.g.*, geometry image and surface normal map [Kakadiaris *et al.* 2007a], shape index image [Huang *et al.* 2011e], signed shape difference map (SSDM) [Wang *et al.* 2010], specific encodings of facial range image, *i.e.*, eLBP [Huang *et al.* 2012] *etc.*. As human faces are all similar each other, elaborated facial representations tend to highlight the distinctiveness of human faces and thereby facilitate the extraction of facial features at the latter stage. A key question here is whether 3D face scans need to be registered before the generation of a proper facial representation. The registration of a 3D face scan consists of aligning it to a common coordinate system so that the matching of facial features can be carried out in a consistent way. As a matter of fact, the generation of a range image, *i.e.* depth map, from a 3D face scan is pose dependent. The net consequence is that all facial representations based on facial range images need first to register properly the raw 3D face scans before their generation.

Popular methods for the registration of 3D face scans are ICP-based [Besl & McKay 1992] which consists of minimizing in an iterative way the distance of two 3D point clouds representing two different 3D face scans. Its shortcomings are well known. The ICP algorithm was designed for the matching of rigid surfaces whereas facial surfaces typically depict elastic deformations. ICP is also quite expensive for aligning 3D face scans of vertices numbered in terms of multiple ten thousands. Besides, ICP-based registration requires in practice that 3D face scans be firstly coarsely registered before their alignment, using for instance a few number of anthropometric points or landmarks located on those 3D face scans, *e.g.*, nose tip, eye corners. Another school of registration methods is called self-dependent alignment [Wang *et al.* 2006c] which consists of aligning 3D face scans through an intrinsic coordinate system of the face defined by 3D landmark structures, *e.g.*, nose tip, nose ridge, symmetry plane [Wang *et al.* 2006c] [Spreeuwers 2011]. However, automatic and accurate localization of 3D landmarks as required in both the two approaches is problematic on partial faces captured in a unconstrained setting [Zhao *et al.* 2011] [Passalis *et al.* 2011].

4.2.2 Facial Features

Given a facial representation, a number of facial features has been proposed in the literature to characterize facial surfaces for the latter stage of matching, including raw face coordinate-based [Maurer *et al.* 2005] [Amor *et al.* 2006c], facial curve-based [Samir *et al.* 2006] [Drira *et al.* 2013], shape descriptor-based [Chua *et al.* 2000] [Li *et al.* 2011c] [Smeets *et al.* 2013], subspace-based [Russ *et al.* 2006] [Spreeuwers 2011], facial representation specific features [Kakadiaris *et al.* 2007a] [Wang *et al.* 2006c] [Huang *et al.* 2012] [Li *et al.* 2011a], *etc.* As a typical pattern recognition problem, good facial features for 3D FR should have the following properties [Huang *et al.* 2012]: first, they should depict high discriminative power to facilitate the latter stage of classification while accounting for large intra-class variations, *e.g.*, pose and expression changes; second, they should be easily extracted from raw facial data to allow fast processing; finally, they should be compact enough so that the matching process can be cost effective and thereby make possible large scale person identification applications.

Facial features can thus be first judged according to the fact whether they are pose-invariant or pose-aligned based [Pears *et al.* 2010]. Most facial features so far proposed in the literature, in particular those extracted from range images, are pose-aligned features. One notable exception is shape-based features, *e.g.*, curvature, shape index, *etc.* In a unconstrained setting, pose-invariant facial features are far preferable than pose-aligned features because 3D face scans can be in an arbitrary pose and face alignment is not only computationally expensive but even problematic in the presence of missing data due to pose and external occlusions.

Another interesting property of facial features is that, while staying discriminating enough for distinguishing different subjects, they could be invariant to variations due to facial expressions. An original essay was made by Bronstein *et al.* [Bronstein *et al.* 2007] who assumed that facial expressions can be modeled as isometries of the facial surface and proposed a facial expression invariant canonical form. However, their assumption proves to be inexact, especially in the presence of exaggerated facial expression [Drira *et al.* 2013].

A far more popular approach observes that facial expressions introduce facial distortions but there are still relatively stable facial regions, *e.g.*, forehead, nose region, from which expression robust features can be extracted [Amor *et al.* 2006b] [Ben Amor *et al.* 2008]. Chang *et al.* [Chang *et al.* 2006] selected three regions around the nose for 3D face matching whereas Faltemier *et al.* [Faltemier *et al.* 2008a] extended the later number to 28 small regions on the face. However, automatic detection and segmentation of facial surface into rigid and mimic regions is still problematic [Amor *et al.* 2006a] [Wang *et al.* 2006c].

Facial features can also be evaluated according to their compactness which impacts the scalability and thereby the applicability of the proposed 3D FR algorithm in real-life biometric applications. From this viewpoint, raw 3D face data, *i.e.* point clouds, which could be used directly as facial features as in [Maurer *et al.* 2005] [Amor *et al.* 2006c] [Mian *et al.* 2007] [Faltemier *et al.* 2008a], are not desirable although the extraction of such a facial feature is obviously straightforward and effortless. Furthermore, the matching of two 3D point clouds generally is carried out through a surface distance, *e.g.*, Mean Square Error (MSE), using ICP which is quite time-consuming, thereby makes the direct use of raw 3D face data as facial features impracticable for identification scenario which requires one-to-many comparisons. Besides, raw 3D face data as facial features are also at odds with the respect of privacy policy in biometric applications which wants to prevent the reconstruction of the original biometric data from their template.

4.2.3 Matching Scheme

3D FR algorithms can also be categorized into three streams according to the nature of their matching scheme: holistic, feature-based or hybrid [Zhao *et al.* 2003] [Mian *et al.* 2008] [Huang *et al.* 2012]. Holistic matching-based 3D FR algorithms perform the matching of global features extracted from the entire 3D face data. Examples of this stream are ICP-based matching [Lu & Jain 2008], subspace-based methods [Heseltine *et al.* 2004], annotated deformable model [Kakadiaris *et al.* 2007a], and isometry invariant description [Bronstein *et al.* 2007]. Holistic 3D FR algorithms utilize global facial features that fully capture facial details and thereby offer potentially very high discriminating power. The downside is that they are likely more sensitive to facial expression variations. Furthermore, for a meaningful matching between holistic facial features, they require an accurate registration step with respect to pose changes. As discussed previously, this is particularly problematic in the presence of missing data and external occlusion for 3D face scans captured under unconstrained conditions from uncooperative subjects.

Feature-based 3D FR algorithms perform the matching of local features extracted from the neighborhoods of a set of keypoints, *e.g.*, feature points automatically detected [Mian *et al.* 2008] [Smeets *et al.* 2009] [Huang *et al.* 2011e] [Li *et al.* 2011c] [Veltkamp *et al.* 2011] or anthropometric landmarks manually or automatically labelled [Berretti *et al.* 2010a]. Much of these works was inspired by the SIFT methodology [Lowe 2004] which was first applied to 2D FR [Bicego *et al.* 2006], then extended to 3D FR in the wave of its extension to 3D range images [Mian *et al.* 2008] or 3D meshes [Russ *et al.* 2009]. As long as the underlying facial representation from which local features are extracted is pose invariant, feature-based matching, as pointed out firstly by Mian *et al.* in 3D FR [Mian *et al.* 2008], has the potential advantage of being robust to facial expression, pose and lighting changes and even to partial occlusions. The downside of

this scheme is the difficulty in extracting sufficient informative feature points from similar or smooth 3D facial surfaces.

Some studies also exist which make use of a hybrid matching in combining global facial features with local ones. Examples of this stream are the component and morphable model-based method [Weyrauch *et al.* 2004], multiple region-based matching [Mian *et al.* 2007], [Faltemier *et al.* 2008a], [Ben Amor *et al.* 2008], [Li *et al.* 2011a]. Most of them consist of utilizing multiple face components, *e.g.*, nose, forehead, which are independently matched through ICP and then fused for a final matching score. As this kind of matching scheme tends to combine facial configuration information with local properties of faces, it is theoretically the most powerful. However, it also risks inheriting both types of shortcomings: necessity of expensive face registration, sensitivity to pose changes and occlusions, difficulty of generating sufficient stable descriptive features, *etc.*

In this chapter, we are concerned with 3D FR in uncontrolled conditions and need to cope with the various challenges, i.e. expression variations, pose changes and external occlusions, at the same time. As a result, we propose a registration free local feature oriented 3D FR method which performs the matching of sparse sets of pose invariant local shape descriptors that we introduce in the following sections.

4.3 Estimating Multiple Order Facial Surface Differential Quantities

In this section, we introduce the estimators of multiple order facial surface differential quantities, which encompass a gradient estimator, a curvature estimator and a derivative of curvature estimator.

4.3.1 Facial Surface Gradient Estimator

Let T be a discrete facial surface represented as a triangular mesh of $\mathbb{E}^3$. T can be regarded as an undirected graph $G=(V,E)$, where $V=\{v_i\}_{1\leq i\leq N}$ and $E=\{e_{ij}\}$ denote the set of vertices and the set of edges respectively. $\forall v_i \in V$, we define $N_1(v_i)=\{v_{i,j}\}_{j=1}^{k}$ as its k adjacent vertices, *i.e.* its one-ring neighbors (Fig. 4.3-a, $i=1,k=5$). We denote $\mathbf{n}_{v_i}$ as the unit normal vector of T and $\mathbf{v}_i$ as the coordinate vector at vertex v_i. Then, $\mathbf{n}_{v_i}$ can be computed as the average of the normals of the triangles incident to v_i (Fig. 4.3-b) as follows:

$$\mathbf{n}_{v_i}=(n_{v_i}^x,n_{v_i}^y,n_{v_i}^z)^{\mathrm{T}}=\frac{1}{k}\sum_{i=1}^{k}\frac{[\mathbf{v}_{i,j}-\mathbf{v}_i]\times[\mathbf{v}_{i,j+1}-\mathbf{v}_i]}{|||[\mathbf{v}_{i,j}-\mathbf{v}_i]\times[\mathbf{v}_{i,j+1}-\mathbf{v}_i]|||} \tag{4.1}$$

Suppose that the local shape of T at vertex v_i can be fitted by a cubic polynomial function $z(x,y)$, then its normal at v_i equals to $(z_x,z_y,-1)^{\mathrm{T}}$. If we rewrite $(n_{v_i}^x,n_{v_i}^y,n_{v_i}^z)^{\mathrm{T}}$ as $(-\frac{n_{v_i}^x}{n_{v_i}^z},-\frac{n_{v_i}^y}{n_{v_i}^z},-1)^{\mathrm{T}}$, the surface gradient direction θ and magnitude can be then estimated as follows:

$$\theta=arctan(\frac{z_y}{z_x})=arctan(\frac{n_{v_i}^y}{n_{v_i}^x}) \tag{4.2}$$

$$||\nabla z(x,y)||=\sqrt{{z_x}^2+{z_y}^2}=\frac{\sqrt{(n_{v_i}^x)^2+(n_{v_i}^y)^2}}{|n_{v_i}^z|} \tag{4.3}$$

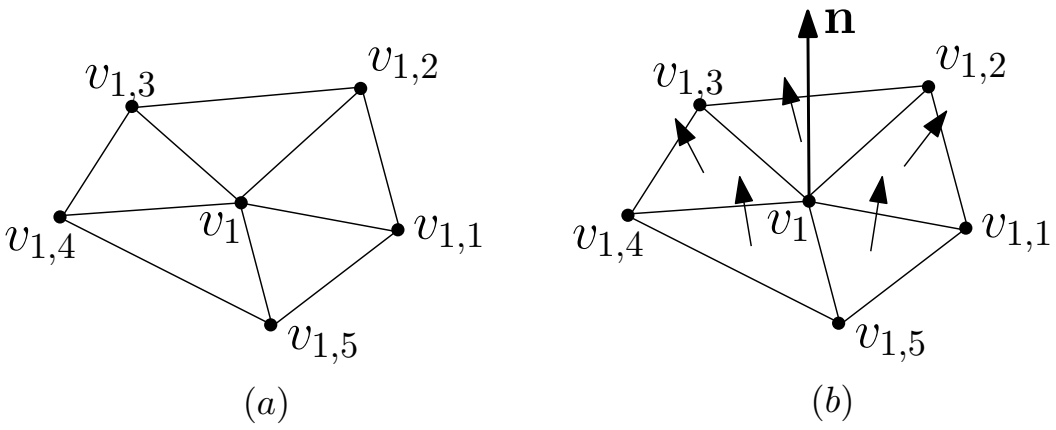


Figure 4.3: *Illustration of normal estimation. (a) one-ring neighborhood, (b)* $\mathbf{n}_{v_1}$ *is computed by averaging the normal vectors of triangles in one-ring of* v_1.

4.3.2 Facial Surface Curvature Estimator

There are three main approaches to calculate curvatures on triangular meshes [GATZKE & Grimm 2006]: a. local fitting; b. discrete estimation of curvature directly on triangular meshes; c. estimation of curvature tensor. In this paper, we make use of the local cubic-order surface fitting method as in [Goldfeather & Interrante 2004b] which displays a better behaviour than discrete estimation method for facial expression analysis [Wang *et al.* 2006a].

For each vertex v_i of T, this method first defines a local 3D coordinate frame with its origin at the vertex and z axis along the normal vector of the vertex. Given two orthogonal axes, x and y, randomly chosen in the tangent plane perpendicular to the normal vector, vertices in the neighborhood (two-ring) of the vertex v_i are then transformed and rotated to the local coordinate system, in which a cubic polynomial function:

$$z(x,y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2 + Dx^3 + Ex^2y + Fxy^2 + Gy^3 \tag{4.4}$$

is approximated by the coordinates of the vertices within the local neighborhood. And it's normal:

$$(z_x, z_y, -1) = (Ax + By + 3Dx^2 + 2Exy + Fy^2, \\ Bx + Cy + Ex^2 + 2Fxy + 3Gy^2, -1) \tag{4.5}$$

is approximated by the normal vectors of the vertices within the local neighborhood. By using least-square fitting method to solve appriximation equations (4.4) and (4.5), we can obtain all the parameters of the local fitting function $z(x,y)$. Then the Weingarten curvature matrix at vertex v_i can be estimated as

$$W_{v_i} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}.$$

The maximum curvature κ_{max} and minimum curvature κ_{min} ($\kappa_{max} \geq \kappa_{min}$) can be estimated as the eigenvalues of W_{v_i}. The shape index can be estimated as:

$$\mathbf{S}(v_i) = \frac{1}{2} - \frac{1}{\pi} arctan(\frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}}) \tag{4.6}$$

4.3.3 Facial Curvature Derivation Estimator

3D FR algorithms require a comprehensive description of shape variations of facial surfaces. As a result, instead of derivatives of deriving curvatures, we compute the gradient of shape index which is a third order differential quantity but more importantly a quantitative measurement of the shape variations of a surface. As in [Russ *et al.* 2009], we define shape index as a scalar function $f : T \rightarrow [0,1]$. Then the gradient of this scalar function f at v_i defined in its local tangent plane can be approximated by solving an optimization problem using the finite element numerical algorithm [Meyer *et al.* 2001].

Let $v_i \in N_1(v_i)$, the gradient $\nabla f(v_i)$ can be estimated by minimizing:

$$\nabla \tilde{f}(v_i) = \arg\min \sum_{j=1}^{k} |\nabla f(v_i)^{\mathrm{T}} \mathbf{P}(\overrightarrow{v_i v_{i,j}}) - \frac{f(v_{i,j}) - f(v_i)}{||\mathbf{v}_{i,j} - \mathbf{v}_i||}| \tag{4.7}$$

where $\mathbf{P}(\overrightarrow{v_i v_{i,j}})$ is the projected unit vector of $\overrightarrow{v_i v_{i,j}}$ in the local tangent plane of v_i. Fig. 4.4 shows the estimated normal vector (b), maximum curvature (c), minimal curvature (d), shape index (e) and gradient of shape index (f) on a sampled facial surface (a).

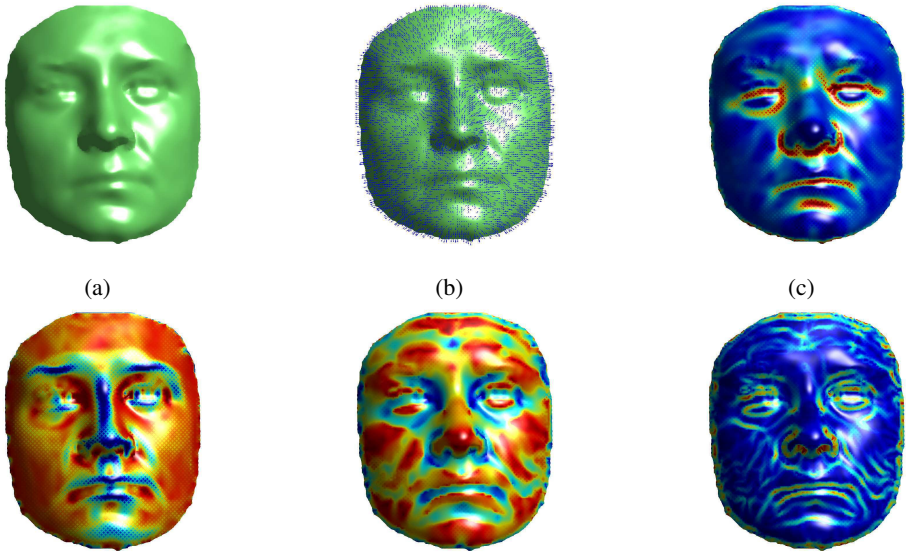


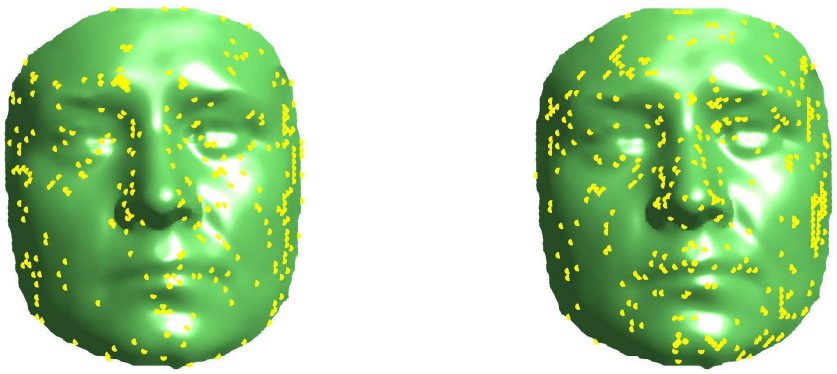
Figure 4.4: *Illustration of multiple order surface differential quantities estimation. (a) original facial surface; (b) normal vector; (c) maximum principle curvature; (d) minimum principle curvature; (e) shape index; (f) gradient of shape index;*

4.4 3D Facial Salient Vertices Detection and their Local Shape Description

Given an input 3D face scan, the purpose of this section is to generate a set of compact yet comprehensive pose invariant local features directly computed over the input 3D mesh to enable the later stage of the matching of partial 3D face scans without any registration. We first describe the automatic detection of facial salient vertices then introduce the pose invariant local shape description of those salient vertices.

4.4.1 3D Facial Salient Vertices Detection: κ_{max} and κ_{min} based Detectors

The detection of salient vertices over 3D face meshes is an extension of the SIFT methodology to 3D [Zaharescu *et al.* 2009] [Maes *et al.* 2010] [Li *et al.* 2011c]. In the case of 3D facial surfaces, the aim here is to automatically locate in the 3D Gaussian scale space salient facial vertices which are evenly distributed over the entire face scan, rigid and elastic regions included, to enable a comprehensive shape description of an input 3D face scan. Specifically, an input raw 3D mesh is firstly smoothed by a Gaussian filter with different values of standard deviation σ. Curvature is then estimated for each vertex on the original and smoothed meshes. Given a vertex v_i, we compute the difference of estimated curvatures of the same vertex for each pair of adjacent scales. Finally, this difference of curvature at a vertex on a given scale is compared to the differences of its one-ring neighbors on its own scale as well as the corresponding ones on upper and lower scales. A vertex is selected as a salient vertex only if its value is a local extreme value within one-ring scale space neighborhood. In order to accurately characterize local shape variations at a given vertex, we propose to make use of maximum curvature (κ_{max}) and minimum curvature (κ_{min}) in this work although their average value, *i.e.* mean curvature, could be another alternative but less precise. Fig. 4.5 shows the detected salient vertices by κ_{max} and κ_{min} detectors, respectively on a sampled facial surface.



(a) (b)

Figure 4.5: *Illustration of the salient vertex detection: salient vertices detected by κ_{max} detector (a), and κ_{min} detector (b), over a facial surface.*

As we can see from Fig. 4.5, salient vertices are evenly distributed over the entire

face surface, containing rigid and elastic facial regions. Furthermore, κ_{max} and κ_{min} detectors locate differently salient vertices which provide added value in 3D FR as we will demonstrate in the experiments.

4.4.2 Pose-invariant Local Shape Description: Histograms of Multi-order Surface Differential Quantities

Once a salient vertex detected on an input 3D face mesh, we introduce several pose-invariant local shape descriptors to comprehensively describe the local shape variations at the detected salient vertex using the histograms of multiple order surface differential quantities. To make the resulted local shape descriptors robust to small transformations, *e.g.*, spatial shifting and non rigid deformations, those histograms are obtained in pooling the multiple order surface differential quantities computed over a properly chosen neighborhood of the detected salient vertex. Specifically, these pose invariant local shape descriptors are computed as follows.

i) Canonical direction(s) assignment. Similar to the original 2D SIFT descriptor, we need to assign a canonical direction(s) to generate a rotation invariant local shape descriptor. Based on the canonical direction(s), two local surface pathes of two different salient vertices can be aligned to be better compared. As shown in Fig. 4.6-a, given a detected salient vertex v (red color), a local geodesic disk surface patch centered at v and with a radius R is considered. Using the same scheme as in section 4.3.2, we first transform the local patch vertices (blue color) to the local coordinate system, where the salient vertex v is the origin, and its normal vector is along the positive z axis. Then, we compute the normal vector of each vertex (Fig. 4.6-b). All these vectors are then projected to the local tangent plane of v, and an initial reference direction is randomly selected (Fig. 4.6-c). Finally, one or several canonical orientations are assigned according to the dominant direction(s) of the projected normal vectors with 360 bins based on the reference direction (Fig. 4.6-d).

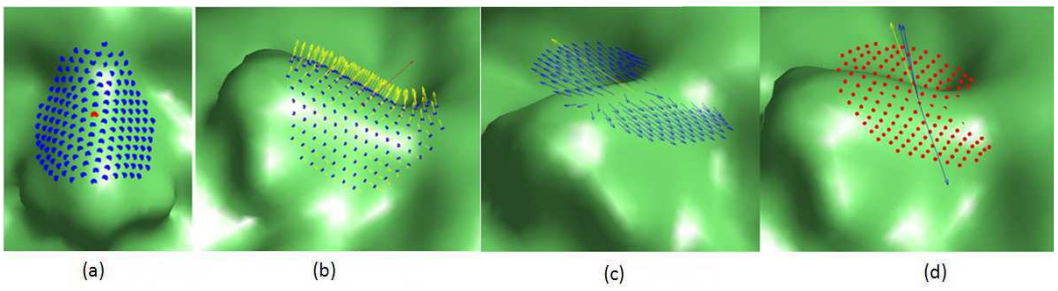


Figure 4.6: *Canonical direction(s) assignment: (a) a detected salient vertex and its geodesic disk patch; (b) coordinate transformation and normal vector estimation; (c) normal vector projection and initial direction selection; (d) angle statistics and canonical direction(s) assignment.*

ii) The configuration of a local shape descriptor. The local shape descriptors associated with a salient vertex are computed in simulating the operation of human complex cells in the visual cortex [Hubel & Wiesel. 1962] so that they are robust to small transformations, *e.g.*, spatial shifting, non rigid deformations. This is achieved in pooling multiple order surface differential quantities computed over the neighborhood of a salient vertex as shown

in Fig. 4.7. Once the canonical orientations are assigned, the local coordinate system rotates in the local tangent plane, making each canonical orientation as new x axis. New y axis can be computed by the cross product of z and x. Based on this new local coordinate system, all the neighbor vertices of a salient vertex are projected to the tangent plane of that salient vertex. Eight projected vertices along to eight quantized directions starting from the canonical orientation with a distance of r_1 to the salient vertex are fixed. 9 circles with a radius r_2, centered at the salient vertex and its eight neighbors, respectively, can be further located. Each of those 9 circles defines a local region which overlaps each other to achieve smooth transitions between those local regions.

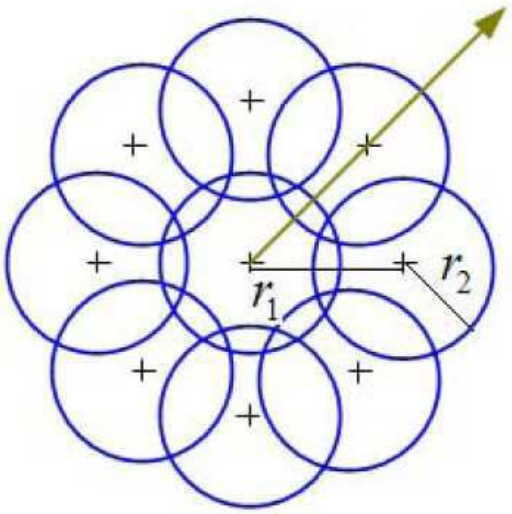


Figure 4.7: *The configuration of a local shape descriptor : Canonical orientation (red vector stemming from the centre vertex with an arrow), the centre salient vertex (+) and its 8 neighborhood vertices (+) with a distance of r_1; 9 circles with the same radius of r_2 are assigned to the 9 vertices, respectively.*

iii) Multiple order surface differential quantities statistics. In each circle i, we pool multiple surface differential quantities and calculate three histograms, namely the histogram of surface gradient (hog_i), histogram of shape index (hos_i) and histogram of gradient of shape index ($hogs_i$). For hog_i and $hogs_i$, the histogram is computed as the gradient angle weighted by gradient magnitude. This histogram is with 8 bins representing 8 main orientations ranging form 0° to 360°. For hos_i, the values of shape index belong to $[0,1]$ are also quantized to 8 bins. Then, all the values of histograms are weighted by a Gaussian function, where the standard deviation is computed as the Euclidian distance between the current vertex and the center vertex of the circle. The final histograms: HOG, HOS and HOGS at a detected salient vertex are formed by concatenating hog_i, hos_i and $hogs_i$ with clockwise direction, represented as follows:

$$\text{HOG} = (hog_1, hog_2, \ldots, hog_9), \tag{4.8}$$

$$\text{HOS} = (hos_1, hos_2, \ldots, hos_9), \tag{4.9}$$

$$\text{HOGS} = (hogs_1, hogs_2, \ldots, hogs_9). \tag{4.10}$$

These three local shape descriptors provide different local shape information within the vicinity of a given vertex. HOG characterizes the orientations where a local surface changes the most and thereby coarsely describe the shape of that local shape. HOS comprehensively describes the various shapes in the vicinity of the given vertex whereas HOGS delivers the speed of local shape changes. We simply concatenate HOG, HOS and HOGS as a single feature vector of a detected salient vertex, and name it as Histogram of Multiple surface differential Quantities (HOMQ):

$$\text{HOMQ} = (hog_1 hos_1 hogs_1, hog_2 hos_2 hogs_2, \ldots, hog_9 hos_9 hogs_9). \tag{4.11}$$

To avoid too big disparities of multi-order surface differential quantities which may typically result from the irregularities of the underlying triangular 3D mesh, all those histograms, *i.e.* hog_i, hos_i, $hogs_i$, $hog_i hos_i hogs_i$, HOG, HOS, HOGS and HOMQ, are further normalized to an unit vector and delivered as local feature vectors. These local feature vectors comprehensively describe local shape variations around a given salient vertex on the input 3D face scan. They are pose invariant and remain very compact. The dimensions of hog_i, hos_i, $hogs_i$ and $hog_i hos_i hogs_i$ equal to 8, 8, 8 and 24, respectively. As a result, the dimensions of HOG, HOS, HOGS and HOMQ are 72, 72, 72 and 216, respectively.

Given a facial triangular mesh T representing a 3D face scan. Let M be the number of salient vertices detected on it and m the total number of canonical directions assigned ($M <= m$). The 3D face scan T can thus be represented by a set of multiple differential quantities based local shape descriptors associated with a set of salient vertices, denoted as,

$$\mathbf{F}_T = \{\mathbf{f_1}, \mathbf{f_2}, \ldots, \mathbf{f_m}\},$$

where $\mathbf{f_i}$, $i \in \{1, 2, \ldots, m\}$ can be HOG, HOS, HOGS or HOMQ. Each of those descriptors is pose invariant, compact and contributes to describe the local shape variations around a given salient vertex. However, depending on the possible missing data due to the face pose and/or external occlusions, the number of those descriptors can vary from one scan to another. In the subsequent, the set of descriptors of all salient vertices detected on the entire face scan are referred to as Multiple salient Verices and multiple Histograms, denoted as MV-HOG, MV-HOS, MV-HOGS and MV-HOMQ, respectively.

4.5 Facial Surface Matching: a SRC-based Fine Grained Matcher vs. a Coarse Grained Matcher

Given two sets of local shape descriptors extracted from two 3D face scans as explained in the previous section, one needs to design a proper matching scheme so as to compare the two input 3D face scans and deliver a similarity measurement. In this section, we introduce a fine grained local feature matching scheme using the Sparse Representation based Classifier (SRC) [Wright *et al.* 2009a] which is put in comparison with a coarse grained local feature matching scheme that was used in our preliminary work [Li *et al.* 2011c]. The coarse grained local feature matching between two face scans, *i.e.* two sets of local shape descriptors, is much inspired by the SIFT methodology in 2D [Lowe 2004] and consists of counting the number of matching pairs of vertices as matching score. However, this is problematic when the underlying face scans have missing data due to arbitrary pose and/or

severe occlusions, which result in a significant decrease in the number of matching pairs of salient vertices. As a result, we propose in this paper a fine grained local feature matching scheme which better accounts for the challenges of missing data and occlusions for 3D FR under uncontrolled setting. It makes use of SRC to compute a mean reconstruction error when each local shape descriptor from a probe face scan is sparsely reconstructed from an over-complete dictionary composed of the local shape descriptors extracted from all the face scans in the gallery. In the subsequent, the coarse grained matcher is first briefly described and then the fine grained matcher is introduced.

4.5.1 The Coarse Grained Matcher (CGM)

Given two facial triangular meshes T_1 and T_2 with m and n salient vertices respectively, we can then extract two sets of local shape descriptors to describe them as explained in section 3:

$$\mathbf{F}_{T_1} = \{\mathbf{f_1^1}, \mathbf{f_2^1}, \cdots, \mathbf{f_m^1}\}$$

and

$$\mathbf{F}_{T_2} = \{\mathbf{f_1^2}, \mathbf{f_2^2}, \cdots, \mathbf{f_n^2}\}.$$

Similar to the SIFT-like matcher which uses the angle of two feature vectors as similarity measurement, we compute $a_{ij} = \arccos\langle \mathbf{f_i^1}, \mathbf{f_j^2} \rangle$, with $i \in \{1,2,\ldots m\}$, and $j \in \{1,2,\ldots n\}$, and define a similarity matrix $[a_{i,j}]$. Each row in that matrix is further ranked in an ascending order. A match between two local shape descriptors is detected if the ratio of the first and the second values in a same row is smaller than a pre-defined threshold μ. The similarity between two 3D face scans T_1 and T_2 is then measured in counting simply the total number of matching pairs. The matching scheme using such a similarity measurement is called *coarse grained matcher* and abbreviated as *CGM* in the subsequent.

This simple coarse grained matching scheme was implemented in several works for 3D FR using facial range images [Mian *et al.* 2008] [Mian *et al.* 2007] [Lo & Siebert 2009] [Zhang & Wang 2009] [Huang *et al.* 2011a] [Huang *et al.* 2012] or 3D facial meshes [Maes *et al.* 2010] [Li *et al.* 2011c] [Smeets *et al.* 2013]. It has been improved to account for the facial configuration in incorporating holistic spatial relationships [Mian *et al.* 2008] [Huang *et al.* 2011a] [Huang *et al.* 2012]. However, its coarse nature in simply counting the number of matching pairs makes it very vulnerable to missing data and external occlusions which typically occur in real-life biometric applications in uncontrolled conditions.

4.5.2 SRC-based Fine Grained Matcher (FGM)

Instead of simply counting the number of matching pairs of salient vertices between two 3D face scans as in CGM, we can estimate a mean reconstruction error when the sparse set of local shape descriptors of the first 3D face scan is used to reconstruct those of the second 3D face scan. The intuition here is that when two 3D face scans are of the same subject and thereby match each other, there should exist, even in case of missing data or occlusions, a set of pairs of salient vertices detected from both the two 3D face scans, which are identical in an ideal case and otherwise very close each other. As a result, when the local shape descriptors of the salient vertices of a first 3D face scan are used

to reconstruct those of the second 3D face scan, the mean reconstruction error should be very small and thereby delivers a fine grained similarity measurement. In this work, we achieve at the same time the identification of matching pairs of local shape descriptors between two 3D face scans and the estimation of mean reconstruction error through a specific sparse representation-based classifier (SRC) firstly introduced by Wright *et al.* [Wright *et al.* 2009a] for 2D FR. In that work, they tried to reconstruct a 2D probe face using those in the gallery and made assumption that a well-aligned probe 2D face image can be represented by a linear combination of gallery 2D facial images of the same subject and thus a sparse linear combination of the whole gallery 2D face images. The reconstruction error of this sparse representation was then used for classification. In this work, SRC is used to reconstruct the local shape descriptors of the salient vertices extracted from an input probe 3D face scan using those extracted from all the gallery faces. Our a priori knowledge here is that a salient vertex of a probe 3D face scan should correspond uniquely to a salient one of the gallery 3D face scan of the same subject as long as that salient vertex is not missing, due to arbitrary pose, or covered, due to external occlusions. In putting into a dictionary the local shape descriptors of all the salient vertices extracted from the gallery 3D face meshes, the local shape descriptor of a given salient vertex on a probe 3D face scan should thus be a very sparse linear combination of those pooled in the dictionary provided that the subject corresponding to the input probe 3D face scan is in the gallery. This idea is implemented here in forming a dictionary with local shape descriptors extracted from all the gallery 3D face scan. The mean reconstruction error of the local shape descriptors extracted from the sparse set of salient vertices in a probe 3D face scan by those of a gallery 3D face scan is then used as a similarity measurement between the probe and that gallery 3D face scan. In comparison with CGM which counts the number of matching pairs of salient vertices, this similarity measurement is a fine grained measure computed in terms of mean reconstruction error. It proves to be more robust to missing data or occlusions and is referred to as *SRC-based Fine Grained Matching scheme* and abbreviated as *FGM* in the subsequent.

4.5.2.1 Building the Gallery Dictionary

Let $T_1, T_2, \ldots, T_C$ be C triangular meshes making up the gallery set of C classes (subjects). Each subject has thus exactly one 3D face scan in the gallery set. Let n_i be the number of salient vertices detected from a 3D face scan T_i, where $i \in \{1, 2, \ldots, C\}$, using a given salient vertex detector, *e.g.*, maximum or minimum curvature-based in this work, the 3D face scan T_i can thus be represented by the set of local shape descriptors extracted from those salient vertices:

$$\mathbf{F}_{T_i} = \{\mathbf{f}_{\mathbf{1}}^{\mathbf{i}}, \mathbf{f}_{\mathbf{2}}^{\mathbf{i}}, \cdots, \mathbf{f}_{\mathbf{n_i}}^{\mathbf{i}}\}.$$

Let $\mathbf{d}_{i,n_i} = \mathbf{f}_{\mathbf{n_i}}^{\mathbf{i}}$ and

$$\mathbf{D}_i = (\mathbf{d}_{i,n_1}, \mathbf{d}_{i,n_2}, \cdots, \mathbf{d}_{i,n_i}).$$

Thus, all the local shape descriptors of class (subject) i are represented by a sub-dictionary $\mathbf{D}_i$ of size $M \times n_i$, where $M = 72$, or $M = 144$, depending on the underlying local shape descriptor, *i.e.* HOG, HOS, HOGS or HOMQ, being used. The overall dictionary for all the C classes in the gallery is thus defined as

$$\mathbf{D} = (\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_C).$$

Note that $\mathbf{D}$ has a total of $K = \sum_{i=1}^{C} n_i$ local shape descriptors, leading to a dictionary of size $M \times K$.

Given a facial surface, hundreds of salient vertices can be detected, making K a large number. That is to say, $\mathbf{D}$ is an over-complete dictionary to represent all the C classes. The local shape descriptor associated with a salient vertex of a probe 3D face scan can thus be sparsely represented using the over-complete dictionary $\mathbf{D}$. Since multiple salient vertices are generally detected from a 3D face scan, multi-task sparse representation is introduced and described in the following subsection.

4.5.2.2 Multi-task Sparse Representation

Given a probe face scan with n salient vertices and n local shape descriptors,

$$\mathbf{Y} = (\mathbf{y_1}, \mathbf{y_2}, \cdots, \mathbf{y_n}), \tag{4.12}$$

the sparse representation of $\mathbf{Y}$ associated with the gallery dictionary $\mathbf{D}$ is formulated as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \sum_{i=1}^{n} \|\mathbf{x_i}\|_0 \text{ s.t. } \mathbf{Y} = \mathbf{DX}, \tag{4.13}$$

where $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}) \in \mathbb{R}^{K \times n}$ is the sparse representation coefficient matrix, and the l_0 norm of vector $\mathbf{x}$ defines the number of its non-zero elements.

$$\|\mathbf{x}\|_0 = \sharp\{i | x_i \neq 0\}.$$

This is a multi-task problem since both $\mathbf{X}$ and $\mathbf{Y}$ have multiple measurements. However, this multi-task problem differs from the one studied in the literature, *e.g.*, joint sparse representation of multiple tasks [Zhang *et al.* 2012], as the observations on 3D face scans here, *i.e.*the local shape descriptors of salient vertices, vary from one 3D face scan to another one, both in terms of number and location. It is thus quite hard to consistently order these salient vertices and capture their collaborative relationships. As a result, we decompose this multi-task problem into n simple sparse representation problems, one for each probe local shape descriptor $\mathbf{y_i}$,

$$\hat{\mathbf{x}}_\mathbf{i} = \arg\min_{\mathbf{x_i}} \|\mathbf{x_i}\|_0 \text{ s.t. } \mathbf{y_i} = \mathbf{Dx_i}, \;\; i = 1, 2, \cdots, n. \tag{4.14}$$

It has been proved that (4.14) is equivalent to solving the following n l_1 minimization problems if the corresponding solution is sparse enough:

$$\hat{\mathbf{x}}_\mathbf{i} = \arg\min_{\mathbf{x_i}} \|\mathbf{x_i}\|_1 \text{ s.t. } \mathbf{y_i} = \mathbf{Dx_i}, \;\; i = 1, 2, \cdots, n. \tag{4.15}$$

There are a number of efficient and fast l_1 minimization algorithms which can be used to solve Eq. (4.15). In this work, we use OMP algorithm [Pati *et al.* 1993] to solve Eq. (4.14), where the sparsity parameter in OMP algorithm is denoted as L. Similar to voting SRC, we use the average reconstruction error to determine the identity of the probe face scan $\mathbf{Y}$.

$$r_j(\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{y_i} - D\delta_j(\hat{x}_i)\|_2^2 \tag{4.16}$$

where δ_j is a characteristic function which selects the coefficients associated with the j-th gallery (sub-dictionary). Finally, the index j which minimizes $r_j(\mathbf{Y})$ in Eq. (4.16) is selected as the identity of $\mathbf{Y}$.

A new matching scheme is thus achieved by solving Eqs. (4.14) and (4.16) and is referred to in the subsequent as SRC-based fine grained matcher (FGM). It is worth to note that this fine grained matching scheme achieves within the same framework not only the matching of the salient vertices of the input 3D probe face scan with those of all the gallery 3D face scans through Eq. (4.15), but also the computation of mean reconstruction errors of the probe 3D face scan with those in the gallery using Eq. (4.16).

4.6 Experimental Results

We evaluated the effectiveness of the proposed approach on Bosphorus, the largest public 3D face database which best accounts for the various challenges of 3D FR in uncontrolled conditions, namely expression variations, pose changes and external occlusions [Savran *et al.* 2008]. As recent studies show that 3D FR algorithms showing state of the art performance over a dataset can vastly degrade on 3D face scans captured in a different scenario, we also carried out additional experiments on FRGC v2.0 [Phillips *et al.* 2005], a popular benchmark in 3D FR, to further check the generalization skill of the proposed approach.

In this section, we first describe in subsection 4.6.1 the Bosphorus database and its various subsets along with a short introduction of FRGC v2.0. We then define in subsection 4.6.2 the experimental setups and protocols and analyze in subsection 4.6.3 the stability of salient vertices located by κ_{max} and κ_{min} detectors and their added value. The experimental results on the the whole Bosphorus and its various subsets are comprehensively analyzed in subsection 4.6.4 in comparison with the state of the art. Subsection 4.6.5 presents the additional experimental results on FRGC v2.0. Finally, we give in subsection 4.6.6some insights of the proposed framework and analyze its time complexity.

4.6.1 Databases

Bosphorus is quite unique for its broad set of expression types, systematic variation of poses and different external occlusions, thus depicts the various challenges that can typically occur in real-life applications under uncontrolled environment. On the other hand, FRGC v2.0 was the first largest dataset of frontal 3D face scans with various facial expressions. In the subsequent, we first comprehensively describe Bosphorus and its various subsets and then introduce shortly FRGC v2.0.

1) The Bosphorus 3D Face Database. The database contains a total number of 4666 3D scans of 105 subjects composed of 61 men and 44 women. For each subject, there are around 34 expressions, 13 poses, and 4 occlusions. 30 professional actors/actresses are incorporated to obtain more realistic expression data. The detailed composition of the three variations are listed as follows:

i) Facial Expressions (FE). This subset contains 6 basic emotions along with neutral, *i.e.* Anger, Disgust, Fear, Happiness, Sadness and Surprise (see Fig. 4.8); 28 facial Action Units (AUs): 20 Lower AUs (LAU), 5 Upper AUs (UAU), 3 Combined AUs (CAU). LAUs occur by the movement of lip(s), mouth, cheek, nose or chin while UAUs correspond to the movement of upper facial components, *e.g.*, eye(s) or brow(s) (see Fig. 4.9 for several

samples). Those facial expressions thus encompass not only prototypical facial expressions but also subtle and exaggerated ones.

Figure 4.8: *2D Samples of 6 basic facial emotions with neutral: (a) neutral; (b) Anger; (c) Disgust; (d) Fear; (e) Happiness; (f) Sadness; (g) Surprise.*

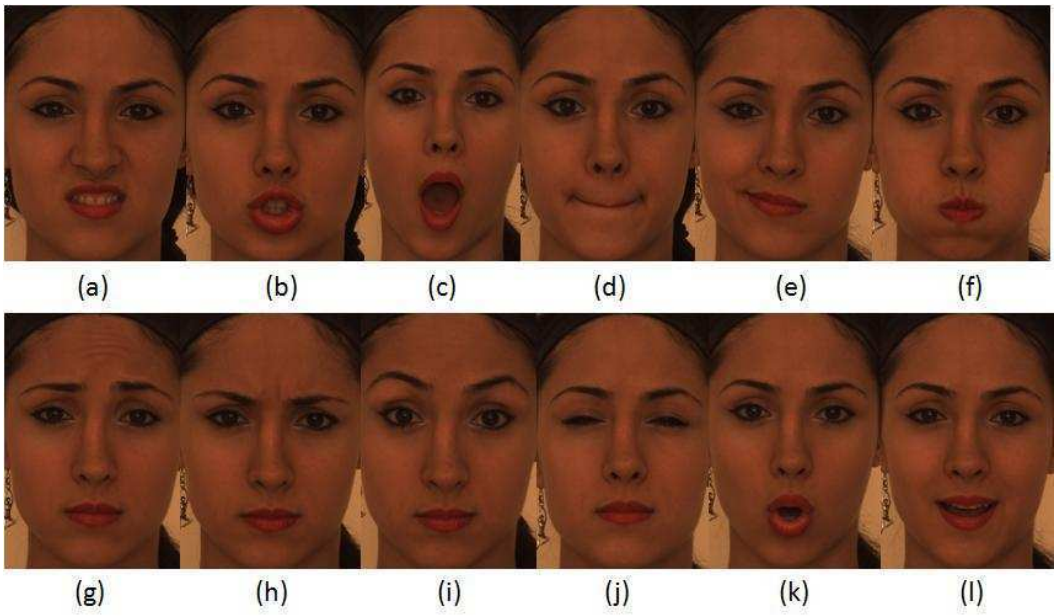


Figure 4.9: *2D Samples of facial action units: (a-f) displaying Lower Action Units, (g-j) displaying Upper Action Units, and (k-l) representing Combined Action Units.*

ii) Poses (P). This subset contains 7 Yaw Rotations (YR): $+10°, +20°, +30°, +45°, +90°, -45°$ *and* $-90°$; 4 Pitch Rotations (PR): Strong upwards/downwards, Slight upwards /downwards; 2 Cross Rotations (CR): $+45°$ yaw and approximately ± 20 pitch (see Fig. 4.10).

iii) Occlusions (O). This subset encompasses: Occlusion of eyes by hand (HandE); Occlusion of mouth by hand (HandM); Occlusion of eyes by glasses (Glasses); Occlusion by hair (long hair of the female and facial hair like beard and moustache of the male) (Hair) (see Fig. 4.11).

vi) A subset of 18 unlabeled frontal scans without occlusion. They are however mostly expressive.

Figure 4.10: *Samples of head pose variations: (a) frontal; (b) Yaw* $+10°$*; (c) Yaw* $+20°$*; (d) Yaw* $+30°$*; (e) Yaw* $+45°$*; (f) Yaw* $-45°$*; (g) Yaw* $+90°$*; (h) Yaw* $-90°$*; (i) PR: Strong upwards; (j) PR: Strong downwards; (k) CR:* $+45°$ *yaw and approximately* $+20$ *pitch; (l) CR:* $+45°$ *yaw and approximately* -20 *pitch.*

Tab. 4.1 shows the distribution of 3D face scans for different subsets. For illustration purposes, only 2D samples are shown in this paper. Please refer to [Savran *et al.* 2008] for more details and more samples of both 2D and 3D scans [1].

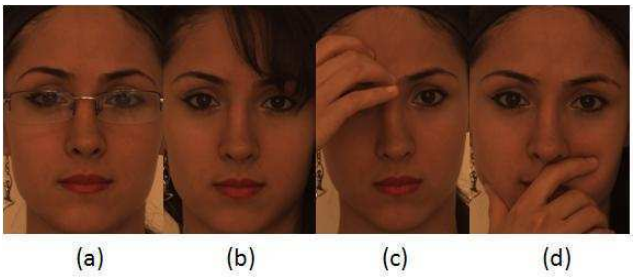


Figure 4.11: *Samples of occlusions: from (a) to (d), Occlusion by glasses, hair, by hand on right eye and by hand on mouth.*

The 3D face scans in that database were acquired using Inspeck Mega Capturor II 3D digitizer device, which has about 0.3mm sensor resolution in all x, y and z dimensions. After the preprocessing steps of manually facial region segmentation, noise removing, and down sampling, each scan approximately contains 35K points.

2) The Face Recognition Grand Challenge v2.0 (FRGC v2.0). The FRGC v2.0 database is the first largest public 3D face dataset with various facial expressions. It consists of 4007 3D scans of 466 subjects acquired using a Minolta Vivid 900 laser scanner. All the scans are (near) frontal and present several facial expression variations, *e.g.*, anger, happiness,

[1] http://bosphorus.ee.boun.edu.tr/default.aspx

Table 4.1: *The distribution of 3D face scans over the various subsets in the Bosphorus database. In the subset of facial expressions, 105 neutral scans, one for each 105 subjects, are used as gallery set whereas the remaining ones as probe sets. The subset YR*45° *of size 210 (YR*90°, resp.*) contains both the subset of 105 scans at Yaw* +45° *(Yaw* +90°, resp.*) and those at Yaw* −45° *(Yaw* −90°, resp.*).*

Expression scans (105+2797)									
Neutral	Anger	Disgust	Fear	Happy	Sad	Surprise	LAU	UAU	CAU
105+194	71	69	70	106	66	71	1549	432	169
Pose scans (1365)									
YR (735)					PR (419)			CR (211)	
YR10°	YR20°	YR30°	YR45°	YR90°	-			-	
105	105	105	210	210	-			-	
Occlusion scans (381)									
HandE			HandM			Glasses		Hair	
105			105			104		67	
unlabeled scans (18)									

sadness, surprise, disgust and cheek puffing. Although FRGC v2.0 only contains frontal 3D face scans captured under controlled lighting conditions, it is a popular benchmark for the assessment of 3D FR algorithms, especially with respect to facial expression variations.

In our experiments, all the scans in Bosphorus were first down sampled and triangulated to triangular meshes by connecting each pair of neighboring vertices. All the scans of FRGCv2 were first preprocessed using the publicly available *3D Face Preprocessing Tools* developed by Szeptycki *et al.* [Szeptycki *et al.* 2009a]. The preprocessing pipeline contains: spike and noise removing, holes filling, nose tip localization, face cropping and triangulation.

4.6.2 Experimental Setups

The proposed approach was first benchmarked on Bosphorus using the identification scenario. Specifically, the first neutral scan of each subject in Bosphorus was used to construct a gallery set and the remaining scans or its subsets were used as probe sets. Therefore, the gallery size is 105 and the probe set size is 4561 for the whole database. Tab. 4.1 gives the various sizes of probe sets tested for each subset.

In the proposed framework, the detection of salient vertices on a 3D face mesh requires the definition of scale parameters σ_i whereas the computation of the local shape descriptors associated with a salient vertex the specification of three radiuses R, r_1 and r_2. In this work, we set the standard deviation of the Gaussian filter $\sigma_i, (i = 1,2,3)$ to 1.83, 2.50 and 4.80, respectively, whereas R, r_1 and r_2 were set to 22.50 mm, 15 mm and 7 mm, respectively. Tab. 4.2 synthesizes these values.

The matching process was carried out using the salient vertices detected separately through the extreme values of κ_{max} and κ_{min} curvatures over the corresponding Gaussian scale space, and their score level fusion was used for final decision. For the coarse grained

Table 4.2: *Parameter Values*

Parameter	σ_1	σ_2	σ_3	R	r_1	r_2
Values	1.83	2.50	4.80	22.50	15	7

matcher (CGM), the number of matched salient vertices based on κ_{max} detector added with that using κ_{min} detector is used for the fused similarity measure for a pair of gallery and probe scans. Similarly, for the SRC-based fine grained matcher (FGM), the sum of the average reconstruction error based on κ_{max} salient vertex detector and that using κ_{min} detector is used for the fused similarity measure for a pair of gallery and probe scans. Tab. G.2 shows the effectiveness of this score level fusion.

For CGM, one needs to set μ, the ratio of the score of the best matching pair over that of the second best matching pair, to an appropriate value. For this purpose, we tested $\mu = 0.70$, $\mu = 0.75$ and $\mu = 0.80$ respectively for different descriptors on the whole Bosphorus database. Based on the results shown in Tab. 4.3, we set $\mu = 0.70$ for HOG and $\mu = 0.75$ for all the other descriptors in all of our experiments. For FGM, in addition to the sparsity parameter L set to 1 as the result of the proposed framework which seeks one-to-one matching of salient vertices, we also consider the setting of $L = 5$ which enables the approximation of a local shape descriptor associated with a salient vertex by those of maximum 5 vertices nearby. The results on the whole Bosphorus dataset using the proposed local shape descriptors are shown in Tab. 4.4. As we can see from that table, these two different settings roughly display the same performance, but the setting with L=1 not only performs much faster but also is our basic assumption in FGM which seeks, for a local shape descriptor of a given salient vertex in a probe face scan, that of the corresponding gallery face scan. As a result, the setting L=1 is adopted in the following experiments.

Table 4.3: *Selection of the parameter μ for the coarse grained matcher (CGM): test on the whole Bosphorus database (105/4561), k_{max} + k_{min} salient vertex detectors.*

%	MV-HOG	MV-HOS	MV-HOGS	MV-HOMQ
$\mu = 0.70$	**82.50**	88.82	80.55	93.33
$\mu = 0.75$	75.18	**90.11**	**81.93**	**94.10**
$\mu = 0.80$	60.75	89.72	81.90	93.77

Table 4.4: *Selection of the sparsity parameter L for the SRC-absed fine grained matcher (FGM): test on the whole Bosphorus database (105/4561), k_{max} + k_{min} detectors.*

%	MV-HOG	MV-HOS	MV-HOGS	MV-HOMQ
$L = 1$	**93.12**	94.37	93.14	**96.56**
$L = 5$	93.05	**94.65**	**93.53**	96.49

Regarding the additional experiments on FRGC v2.0, the popular protocol as used in [Faltemier *et al.* 2008a] [Mian *et al.* 2008] [Kakadiaris *et al.* 2007a] [Huang *et al.* 2012] is adopted. Specifically, the first scans of each subject is used to construct a gallery set whereas the remaining 3D scans or its subsets as probe sets. The gallery size is 466 and the probe set size is 3561 for the whole database. The probe sets of other two subsets are the neutral and non-neutral sets, respectively.

4.6.3 Analysis of Salient Vertices Located by κ_{max} and κ_{min} Detectors

In the proposed framework, each 3D face scan is represented by the set of pose invariant local shape descriptors, namely statistics of multiple order surface differential quantities, computed over a set of salient vertices using κ_{max} and κ_{min} detectors in their respective Gaussian scale space. The stability of detected salient vertices detected, *i.e.* the repeatability and their number over 3D face scans are thus a key factor which is analyzed in-depth in this subsection using Bosphorus. Furthermore, we also highlight the potential added value of these two salient vertices detectors for the purpose of 3D FR.

4.6.3.1 On the Stability of Detected Salient Vertices

Given two arbitrary face scans of a same subject, the repeatability of detected salient vertices implies that the two sets of salient vertices extracted from the two face scans at a same pose are roughly located at the same facial positions. However, it is quite hard to have ground truth to check such a repeatability over 3D face scans of the same subject. In this paper, we adopted the protocol proposed by Mian *et al.* [Mian *et al.* 2008] and checked the repeatability of salient vertices as follows: (i) we considered all face scans in the gallery set as reference scans and registered all probe scans to its corresponding reference scan, *i.e.* the gallery scan of the same subject, using the ICP algorithm [Zhang 1994]; (ii) For each salient vertex of the test scan, we located its closest salient vertex on the reference scan and their distance was recorded as its error. As a result, for all the test scans, one can display the cumulative repeatability percentage as a function of the increasing error tolerant distance. Fig. 4.12 plots such a cumulative repeatability of salient vertices on the 3D face scans of the subset of expression in Bosphorus. From that figure, we can see that: (i) κ_{min} detector is more stable than κ_{max} detector, within a distance of 6 mm, the repeatability of κ_{min} detector reaches 75.3% (67.6%, resp.) while κ_{max} detector achieves 67.62% (61.01%, resp.) for neutral face scans (non neutral face scans, resp.) individually. (ii) Both the repeatability of κ_{min} detector and κ_{max} detector record significant drop from neutral face scans to non neutral ones. This is expected as facial expressions elastically deform facial surfaces and thereby induce different salient vertices being located. However, the repeatability in such case still remains as high as 67.6% for κ_{min} detector (61.01% for κ_{max} detector, resp.), meaning that much local regions around detected salient vertices can still be used for matching.

The proposed 3D FR achieves the description of a 3D face scans using a sparse set of pose invariant local shape descriptors computed around salient vertices. The number and spatial locations of those salient vertices thus impact the quality of such a description. Fig. 4.5 suggests that detected salient vertices are rather evenly distributed over an entire face scan. But how their number is impacted by the various challenges, *i.e.* variations of

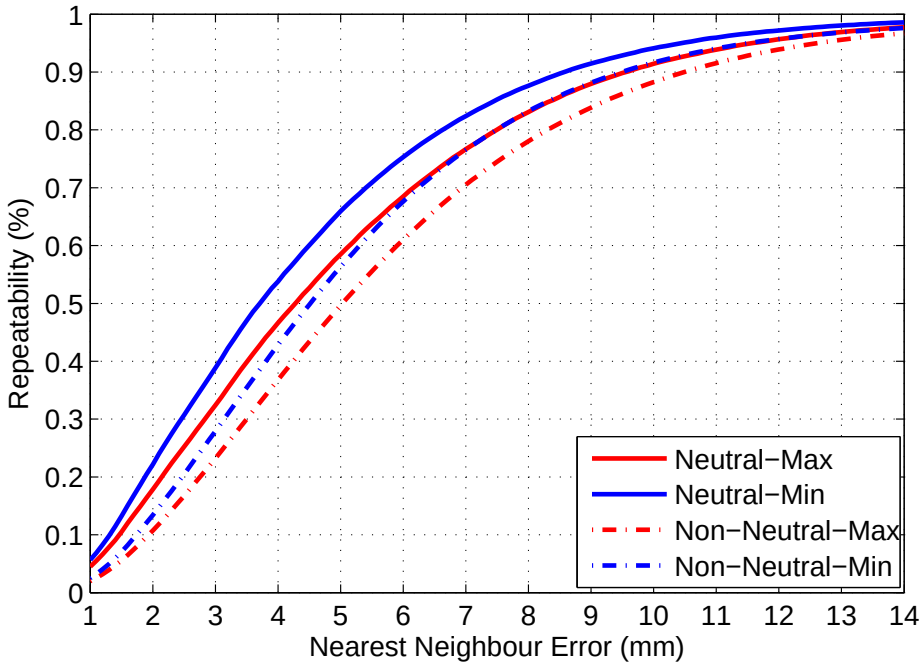


Figure 4.12: *Repeatability of salient vertices on the Bosphorus database.*

expression, pose and occlusion, in 3D FR under uncontrolled conditions? For this purpose, Fig. 4.13 plots the average number of detected salient vertices across different variations of expression, pose and occlusions.

As we can see from Fig. 4.13, the average number of salient vertices detected by both the two κ_{max} and κ_{min} detectors remains relatively stable across different facial expression classes, moderate pose changes up to 20° and occlusions. However, when the pose change significantly increases, *e.g.*,from yaw 20° to 90°, the average number of salient vertices detected by the κ_{max} detector (κ_{min} detector, resp.) also decreases from more than 300 (350, resp.) to less than 200 (250, resp.). However, this decrease rate, roughly 30%, is less rapid than the missing data rate which is of 50% when a face is rotated from a frontal pose to a profile one (e.g., Yaw 90°). Furthermore, we find that external occlusions, *e.g.*, hand, scarf, don't lead to a decrease of detected salient vertices on 3D face scans, as they introduce different surfaces covering the underneath facial surfaces. As we will see later on, this decrease of detected salient vertices will impact the performance of the proposed framework.

Tab. 4.5 further synthesizes the average number of salient vertices located by both the κ_{max} and κ_{min} detectors on the whole Bosphorus and its three subsets of expression, pose and occlusion. As we can see from that table, the κ_{min} detector detects in average more salient vertices than the κ_{max} detector. This can be explained by the concave nature of facial surfaces. Furthermore, when there is no disturbing factor, *e.g.*,pose, leading to significant missing face data, the average number of detected salient vertices remains relatively stable, roughly 300 by the κ_{max} detector and 350 by the κ_{min} detector.

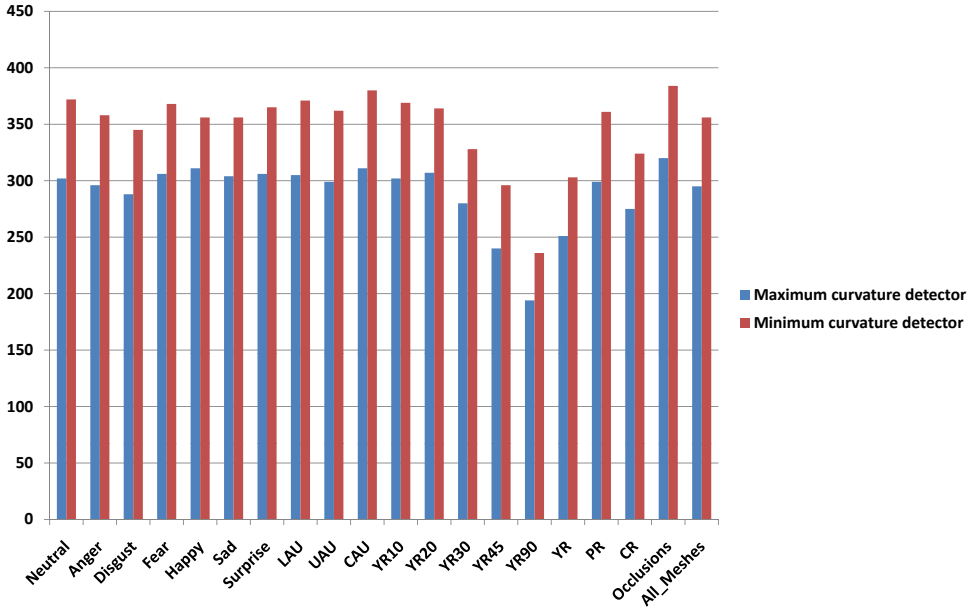


Figure 4.13: *The average number of detected salient points on each subset of the Bosphorus database.*

Table 4.5: *The average number of detected salient vertices on the subsets of expression, pose and occlusion and the whole Bosphorus database.*

	Expression		Pose		Occlusion		All	
	κ_{max}	κ_{min}	κ_{max}	κ_{min}	κ_{max}	κ_{min}	κ_{max}	κ_{min}
Mean	304	368	270	324	320	384	293	355

4.6.3.2 The added value of the joint use of κ_{max} and κ_{min} detectors

A natural question of using both the κ_{max} and κ_{min} detectors is their added value for the purpose of 3D FR. The previous subsection already reports that they locate different number of salient vertices on 3D face scans. In this subsection, we further highlight their complementarity and added value in 3D FR in comparing the joint use of their detected salient vertices through both the proposed coarse grained matcher (CGM) and fine grained matcher (FGM) with the single use of each alone. Tab. G.2 displays the results of this comparison on the whole Bosphorus dataset using the proposed 4 local shape descriptors, namely MV-HOG, MV-HOS, MV-HOGS and MV-HOMQ. As we can see from that table, in comparison with the cases where the salient vertices located by κ_{max} and κ_{min} detectors are used separately, both CGM and FGM significantly improves their performance for each of these local shape descriptors when those salient vertices are jointly used through a score level fusion using

a simple sum rule. This complementarity and added value is further highlighted using the Cumulative Match Curves (CMC) as plotted in Fig. 4.17 (c) and (d).

Table 4.6: *The complementarity and added value of using both k_{max} and k_{min} detectors on the whole Bosphorus database (105/4561).*

%	CGM			FGM		
	κ_{max}	κ_{min}	$\kappa_{max} + \kappa_{min}$	κ_{max}	κ_{min}	$\kappa_{max} + \kappa_{min}$
MV-HOG	69.57	76.21	**82.50**	86.65	90.79	**93.12**
MV-HOS	82.53	85.31	**90.11**	91.80	92.83	**94.37**
MV-HOGS	71.32	75.47	**81.93**	91.87	88.75	**93.14**
MV-HOMQ	87.33	91.01	**94.10**	94.78	95.35	**96.56**

4.6.4 Face Recognition on the Bosphorus Database

We benchmarked the proposed 3D FR method on Bosphorus, the most comprehensive public database which best accounts for the various challenges of real-life applications in uncontrolled conditions, *e.g.*, facial expression variations, pose changes and external occlusions.

4.6.4.1 The Results

Tab. 4.7 reports the rank one recognition rates of the proposed method on the whole Bosphorus database and its various subsets. The CGM is compared with the fine grained matcher over the proposed four local shape descriptors, namely MV-HOG, MV-HOS, MV-HOGS and MV-HOMQ which results from the early fusion, *i.e.* the simple concatenation, of the first three descriptors. Fig. 4.14 plots the performance of CGM over the various subsets of Bosphorus when the aforementioned four local shape descriptors are individually used. Fig. 4.15 does the same for FGM.

Tab. 4.7 can be analyzed from several viewpoints, *e.g.*, the expressiveness of the proposed local shape descriptors, the effectiveness of the fusion schemes, the quality or fineness of the matchers, and the robustness of the proposed framework with respect to the various challenges of 3D FR under uncontrolled conditions. Thus, it can be seen from Tab. 4.7 that:

(i) In terms of descriptive power of single local shape descriptors, MV-HOS performs the best in the class of single local shape descriptors in comparison with MV-HOG and MV-HOGS over both the two CGM and FGM matchers and across all 3D face variations. This is rather expected as HOS computed over the neighborhood of a given vertex is the histogram of shape index values which accurately characterize the local shape of that neighborhood whereas HOGS, *i.e.* the histogram of gradient of shape index, measures the speed of shape variations in the vicinity of a vertex. Finally, HOG is the histogram of surface gradients which delivers the orientations where a local surface change the most and thereby only coarsely describes the geometric shape of a local facial surface.

Table 4.7: *Rank-1 recognition rates with respect to Expression, Pose and Occlusions Variations on the Bosphorus database . CGM is compared with the FGM over the four different local shape descriptors. The meaning and composition of each subset is detailed in Tab. 4.1.*

CGM (left) vs. FGM (right)								
Expression Variations (2797 scans)								
	MV-HOG		MV-HOS		MV-HOGS		MV-HOMQ	
Neutral	99.48	**100.0**	100.0	**100.0**	99.48	**100.0**	100.0	**100.0**
Anger	69.01	**91.55**	87.32	**98.59**	76.06	**94.37**	88.73	**97.18**
Disgust	50.72	**84.06**	56.52	**78.26**	60.87	**88.41**	76.81	**86.96**
Fear	71.43	**92.86**	88.57	**92.86**	84.29	**94.29**	92.86	**98.57**
Happy	79.25	**91.51**	85.85	**96.23**	75.47	**98.11**	95.28	**98.11**
Sad	80.30	**95.45**	90.91	**96.97**	86.36	**96.97**	95.45	**100.0**
Surprise	81.69	**100.0**	97.18	**100.0**	84.51	**97.18**	98.59	**98.59**
LAU	88.96	**96.97**	95.09	**98.13**	90.70	**98.52**	97.22	**98.84**
UAU	94.21	**99.07**	98.15	**100.0**	95.14	**99.54**	99.07	**100.0**
CAU	92.90	**100.0**	97.04	**99.41**	95.86	**100.0**	98.82	**100.0**
Overall	88.09	**96.96**	94.32	**97.96**	90.24	**98.32**	96.89	**98.82**
Pose Variations (1365 scans)								
	MV-HOG		MV-HOS		MV-HOGS		MV-HOMQ	
YR10°	98.10	**100.0**	100.0	**100.0**	96.19	**100.0**	100.0	**100.0**
YR20°	90.48	**99.05**	96.19	**99.05**	87.62	**98.10**	99.05	**100.0**
YR30°	84.76	**98.10**	92.38	**100.0**	75.24	**96.19**	98.10	**99.05**
YR45°	49.52	**89.05**	80.48	**91.90**	49.05	**84.76**	90.95	**97.62**
YR90°	07.62	**16.67**	17.62	**25.24**	02.86	**10.00**	33.33	**47.14**
YR	55.37	**72.65**	69.25	**76.19**	51.84	**71.43**	69.12	**84.08**
PR	94.27	**99.28**	96.90	**98.57**	84.73	**99.28**	98.81	**99.52**
CR	60.66	**90.05**	79.62	**94.79**	48.34	**89.10**	94.31	**99.05**
Overall	68.13	**83.52**	79.34	**85.93**	61.39	**81.47**	86.89	**91.14**
Occlusions (381 scans)								
	MV-HOG		MV-HOS		MV-HOGS		MV-HOMQ	
HandE	96.19	**100.0**	100.0	**100.0**	96.19	**100.0**	100.0	**100.0**
HandM	90.48	**100.0**	97.14	**98.10**	93.33	**99.05**	99.05	**100.0**
Glasses	95.19	**100.0**	97.38	**97.12**	100.0	**97.12**	100.0	**100.0**
Hair	86.57	**94.03**	**94.03**	91.01	**86.57**	82.09	**97.01**	95.52
Overall	92.65	**98.95**	97.38	**97.90**	93.96	**96.59**	99.21	**99.21**
Unlabeled scans (18 scans)								
	MV-HOG		MV-HOS		MV-HOGS		MV-HOMQ	
Overall	88.89	**100.0**	100.0	**100.0**	94.44	**100.0**	100.0	**100.0**
All Meshes Except Yaw 90 (4351 scans)								
	MV-HOG		MV-HOS		MV-HOGS		MV-HOMQ	
Overall	86.12	**96.81**	93.61	**97.70**	85.75	**97.15**	97.04	**98.94**
All Meshes (4561 scans)								
	MV-HOG		MV-HOS		MV-HOGS		MV-HOMQ	
Overall	82.50	**93.12**	90.11	**94.37**	81.93	**93.14**	94.10	**96.56**

(ii) Regarding the effectiveness of the fusion schemes, recall that the proposed framework proposes two different fusion schemes: an early fusion, *i.e.* a simple concatenation of the previous three single local shape descriptors, namely MV-HOG, MV-HOS and MV-HOGS, leading to MV-HOMQ, and a late fusion through the simple sum of scores when both the salient vertices located by κ_{max} and κ_{min} detectors are jointly used. Fig. 4.17 plots the Cumulative Match Curves (CMC) in highlighting the usefulness of these two fusion schemes. In 4.6.3.2, we already discussed the added value of using jointly the salient vertices located by both the κ_{max} and κ_{min} detectors through the score-level fusion. We focus our attention here on the usefulness of feature level fusion as shown in Fig. 4.17 (a) and (b). In capturing the complementary local shape information as provided by MV-HOG, MV-HOS and MV-HOGS, MV-HOMQ indeed achieves a more comprehensive description of local shapes than each of the three single local shape descriptors, and thereby improves their individual performance through both the two matchers CGM and FGM, aside from only the two cases of anger and disgust. This improvement is particularly significant in case of severe missing face data, *e.g.*, YR90°, where the rank-1 recognition rate using CGM (FGM, *resp*.) is increased from 7.62% (16.67%, *resp*.) by MV-HOG, 17.62% (25.24%, *resp*.) by MV-HOS and 2.86% (10%, *resp.*) by MV-HOGS to 33.33% (47.14%, *resp*.) by MV-HOMQ.

(iii) Concerning the matcher quality, the proposed SRC-based fine grained matcher (FGM) clearly demonstrates its effectiveness in outperforming the coarse grained matcher (CGM) over the whole Bosphorus dataset. The amelioration by FGM is even general, across all local shape descriptors and all face variations, aside from the only case of MV-HOMQ in dealing with the subset of external occlusions. This improvement is particularly impressive in hard cases with pose variations, *e.g.*, YR45°,YR90°, and weak descriptors, *e.g.*, MV-HOG. Indeed, in the case of YR45° with MV-HOG, FGM achieves 49.52% rank-1 recognition rate which increases by 40 points the performance of 89.05% by CGM. Fig. 4.16 further highlights such a behavior. In the case of using a more comprehensive local shape descriptor, *i.e.* MV-HOMQ, the improvement by FGM is less impressive but still significant, in particular in the extreme case of using 3D face profiles as probe set, *i.e.* the subset of YR90°. In that case, FGM achieves still 48.57% rank-1 recognition rate and thus improves by roughly 15 points the rank-1 recognition rate (33.33%) achieved by CGM.

(iv) In terms of robustness to the various challenges of 3D FR in uncontrolled conditions, the proposed local feature-based framework displays its effectiveness, in particular when MV-HOMQ, the most comprehensive local shape descriptor, is used. However, the fine grained matcher (FGM), aside from the only case of occlusions, clearly outperforms the coarse grained matcher (CGM), in particular when missing face data due to pose is increasing. Specifically, when dealing with expression variations, the configuration MV-HOMQ/FGM of the proposed framework, *i.e.* MV-HOMQ along with the fine grained matcher FGM, achieves very high recognition rates on all the expression subsets excluding Disgust, and even attains 100% rank-1 recognition rates in dealing with half of them, *i.e.* Neutral, Sad, Surprise, UAU, CAU, respectively. The overall rank-1 recognition rate of the configuration MV-HOMQ/FGM in dealing with the subset of all the 2797 expressive 3D face scans attains 98.75%. Regarding the challenge of pose variations, in excluding the extreme case of YR90 where 3D face scans only contain the profile face, the configuration MV-HOMQ/FGM achieves once more a high rank-1 recognition rate of 98.80% and even 100% rank-1 recognition rates in dealing with the subsets of YR10, YR20 and

YR30, respectively. Finally, in handling the subset of occlusions, the configuration MV-HOMQ/FGM only misclassifies 3 face scans heavily occluded by hair (see Fig. 4.23) and thus displays a rank-1 recognition rate up to 97.90%. All these figures consistently demonstrate the robustness of the proposed framework with the challenges of expressions variations, pose changes and occlusions. Please see Fig. 4.18 for the CMC curves of several sample expression and pose subsets and Fig. 4.19 for the CMC curves of the configurations of MV-HOMQ/CGM and MV-HOMQ/FGM on the whole Bosphorus database.

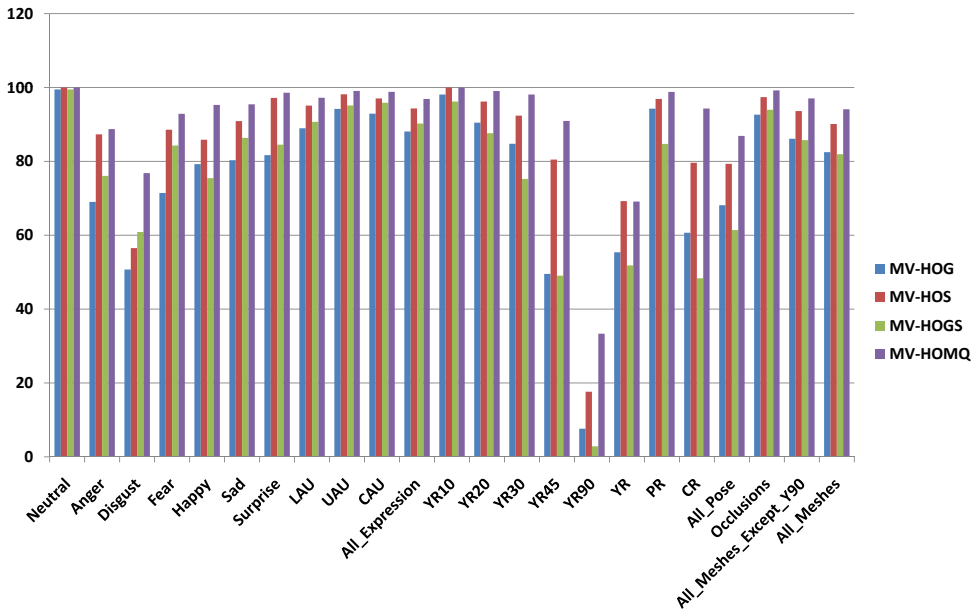


Figure 4.14: *The rank-1 recognition rates by the coarse grained matcher (CGM) over the various subsets of the Bosphorus database.*

4.6.4.2 Comparison with the State of the Art

Tabs. 4.8, 4.9 and 4.11 compare the proposed 3D FR method with the state-of-the-art work on the expression and pose subsets, and the whole set of the Bosphorus database, respectively. Indeed, the Bosphorus database encompasses such a joint level of challenges that most state of the art works are only benchmarked using a subset of Bosphorus in targetting a specific challenge.

Tab. 4.8 compares the proposed 3D FR method with the state of the art in terms of robustness to facial expression variations. As it can be seen, early works, *e.g.*, [Alyuz *et al.* 2008] [Dibeklioğlu *et al.* 2009], only evaluated their methods on a subset of the Bosphorus database. They performed a two phase registration with a first phase of coarse registration which is followed by a second phase of local dense registration through a set of average regional models. A follow-up work of the same group proposes to segment

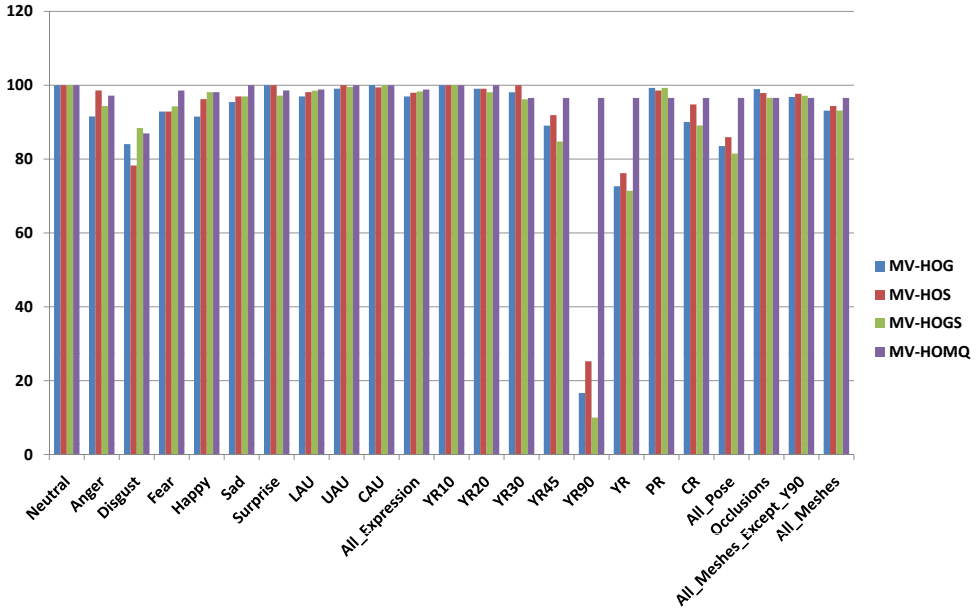


Figure 4.15: *The rank-1 recognition rates by the fine grained matcher (FGM) over the various subsets of the Bosphorus database.*

the nasal region for 3D FR under pose and expression variations where the nose direction is utilized to correct pose variations [Dibeklioğlu *et al.* 2009]. They further improved their method later on in learning discriminating 3D features through LDA and fusing region-based registration classifiers [Alyüz *et al.* 2010]. The LDA-based learning strategy was also adopted by Ocegueda *et al.* [Ocegueda *et al.* 2011a] to select highly discriminating wavelet coefficients extracted from a geometry image, the later being computed after a fine multistage alinement and the fitting of an Annotated Face Model (AFM) to an input probe scan [Kakadiaris *et al.* 2007a]. Maes *et al.* [Maes *et al.* 2010] developed a similar local feature-based approach as compared the proposed method and displayed a rank-1 recognition rate of 97.7% on all the frontal probe scans. As we can see from Tab. 4.8, without any prior registration, the proposed method with the configuration MV-HOMQ/FGM achieves the best rank-1 recognition rates in dealing with both the subset of all frontal scans or the one of all expressive scans.

Tab. 4.9 compares the proposed method with the state of the art on the subset of pose variations in Bosphorus. As we can see from that table, very few methods in the literature tackle this problem and take this hard subset as benchmark. Hajati *et al.* [Hajati *et al.* 2012] reported an approach using patch geodesic moments which are extracted from the texture image controlled by the corresponding range image. However, the cropping, pose correction and alignment in their method were made manually. Comparatively, the proposed method achieves much better results over all subsets of pose changes using the mesh-based 3D scans without any prior information. It is worth to note that, with a rank-1 recognition rate less than 50%, matching profile 3D face scans in the extreme case of YR90 is still

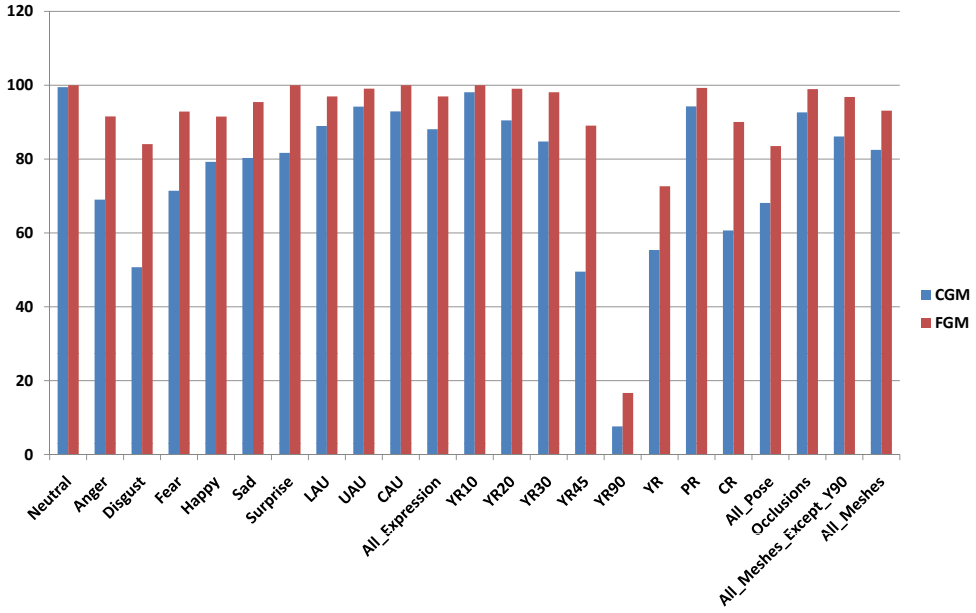


Figure 4.16: *Comparison of the coarse grained matcher (CGM) and the fine grained matcher (FGM) on the whole Bosphorus database when MV-HOG is used as local shape descriptor.*

problematic and needs further investigation.

Tab. 4.10 compares the proposed method with the state of the art with respect to occlusion variations using the corresponding subsets of Bosphorus. Very few methods deal explicitly with this issue. State of the art techniques, *i.e.* [Colombo *et al.* 2011a][Drira *et al.* 2013], are learning-based approach which proposes to detect and restore partially occluded face parts using a subspace of 3D face scans trained with an additional dataset. As such, both training 3D face scan and probe scans need to be correctly registered. Colombo *et al.* [Colombo *et al.* 2011a] benchmarked their method using the probe subset of occlusion reduced to 360 scans by discarding scans judged as of bad quality. Drira *et al.* [Drira *et al.* 2013] made use of the nose tip coordinates provided by the Bosphorus dataset to preprocess and align probe face scans.As we can see from Tab. 4.10, the proposed framework, without any prior training and alignment, outperforms these state of the art methods qui require both learning and registration.

Finally, tab. 4.11 presents a comparison of overall performance on the whole Bosphorus database. To the best of our knowledge, there is only another work in the literature which has the ability of dealing with the whole Bosphorus database. Once more, we can see that our method outperforms that method.

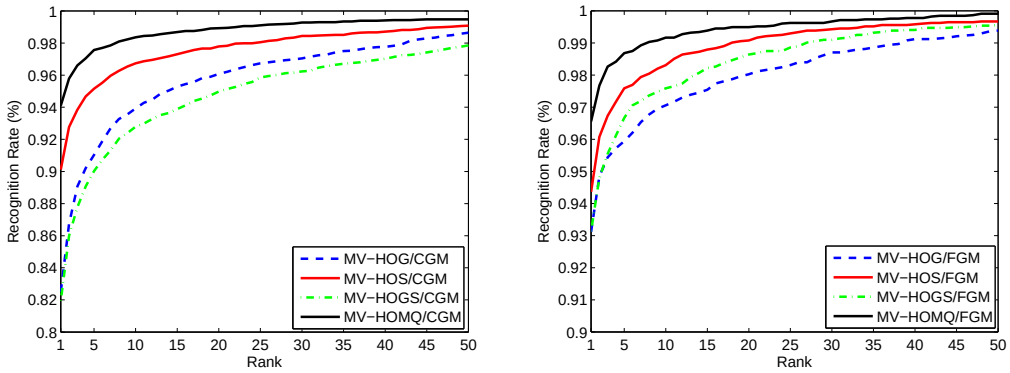


Figure 4.17: *On the usefulness of salient point detector fusion and feature fusion. Early fusion of local shape descriptors: (a) coarse grained matcher (CGM); (b) fine grained matcher (FGM). Score-level fusion of both the κ_{max} and κ_{min} detectors: (c) CGM; (d) FGM.*

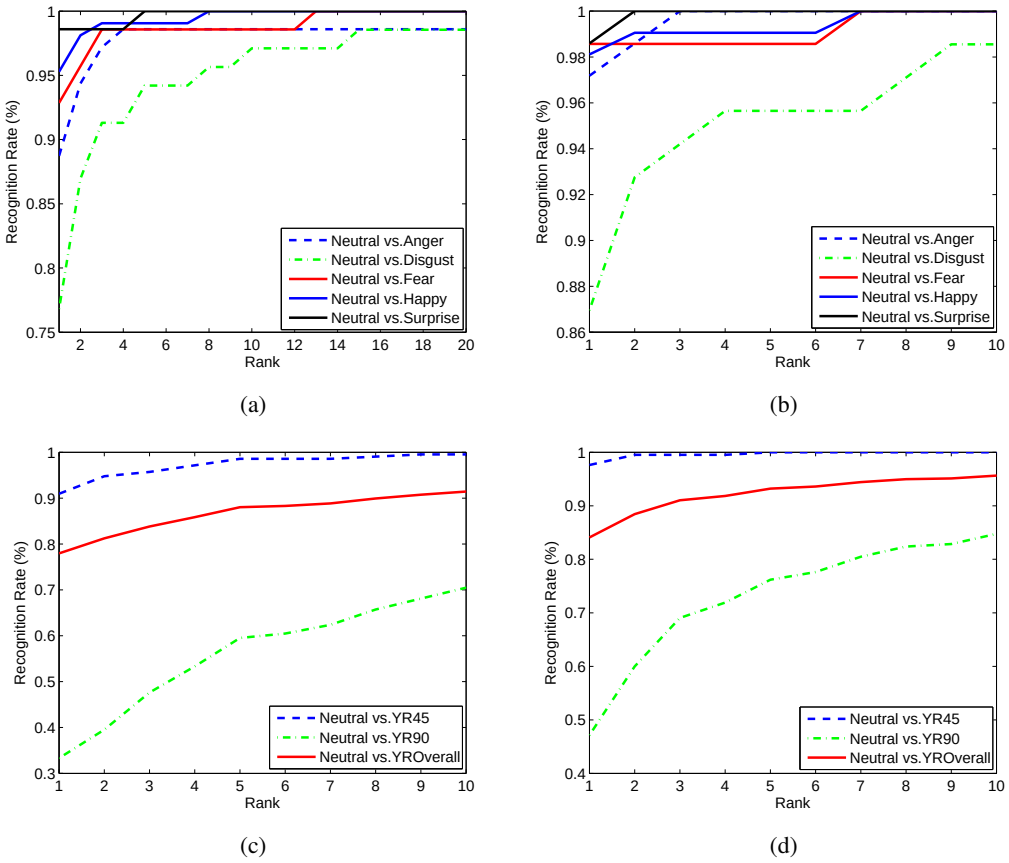


Figure 4.18: *Illustration of performance on expression and pose subsets. (a) performance of coarse-matcher on expression subsets; (b) performance of fine-matcher on expression subsets; (c) performance of coarse-matcher on pose subsets; (d) performance of fine-matcher matcher on pose subsets.*

Table 4.8: *Performance comparison on the subset of expressions in the Bosphorus database. The subset frontal encompasses all probe 3D face scans, the subset of occlusion and unlabeled scans included, and only excludes the subset of pose.*

Approaches	Gallery	Probe	Registration	Rank-one
Baseline ICP [Alyuz *et al.* 2008]	47	1508	Yes	72.41%
Baseline PCA (whole face) [Alyuz *et al.* 2008]	47	1508	Yes	70.56%
Base line PCA (Eye, Nose) [Alyuz *et al.* 2008]	47	1508	Yes	88.79%
Alyuz *et al.* [Alyuz *et al.* 2008]	47	1508	Yes	95.3%
Dibeklioglu *et al.* [Dibeklioglu *et al.* 2009]	47	1527	Yes	89.2%
Alyuz *et al.* [Alyuz *et al.* 2010] (AvRM-PC-LDA)	85	2180	Yes	99.3%
Alyuz *et al.* [Alyuz *et al.* 2010] (AvRM-CD)	105	2814	Yes	98.2%
Kakadiaris *et al.* [Ocegueda *et al.* 2011a] (expression)	105	2797	Yes	98.2%
Maes *et al.* [Maes *et al.* 2010] (frontal)	105	3186[2]	No	97.7%
MV-HOMQ/FGM (expression)	105	2797	No	**98.8%**
MV-HOMQ/FGM (frontal)	105	3196	No	**98.9%**

Table 4.9: *Performance comparison on the subset of the Bosphorus database with various pose variations.*

%	[Dibeklioglu *et al.* 2009]	[Maes *et al.* 2010]	[Hajati *et al.* 2012]	MV-HOMQ/FGM
YR10°	-	-	92.3	**100**
YR20°	-	-	88.6	**100**
YR30°	-	-	80.0	**99.1**
YR45°	-	85.6	38.6	**97.6**
YR90°	-	24.3	-	**47.1**
PR	-	-	-	99.5
CR	-	-	-	99.1
Overall	79.41	84.2	69.1	**91.1**

Table 4.10: *Performance comparison on the subset of the Bosphorus database with diverse occlusions.*

%	Eye	Mouth	Glasses	Hair	Overall
Alyuz *et al.* [Alyuz *et al.* 2008] (381 scans)	93.6	93.6	97.8	89.6	93.6
Colombo *et al.* [Colombo *et al.* 2011a] (360 scans)	91.1	74.7	94.2	90.4	87.6
Drira *et al.*[Drira *et al.* 2013](381 scans)	97.1	78	94.2	81	87
MV-HOMQ/FGM (381 scans)	**100**	**100**	**100**	**95.5**	**99.2**

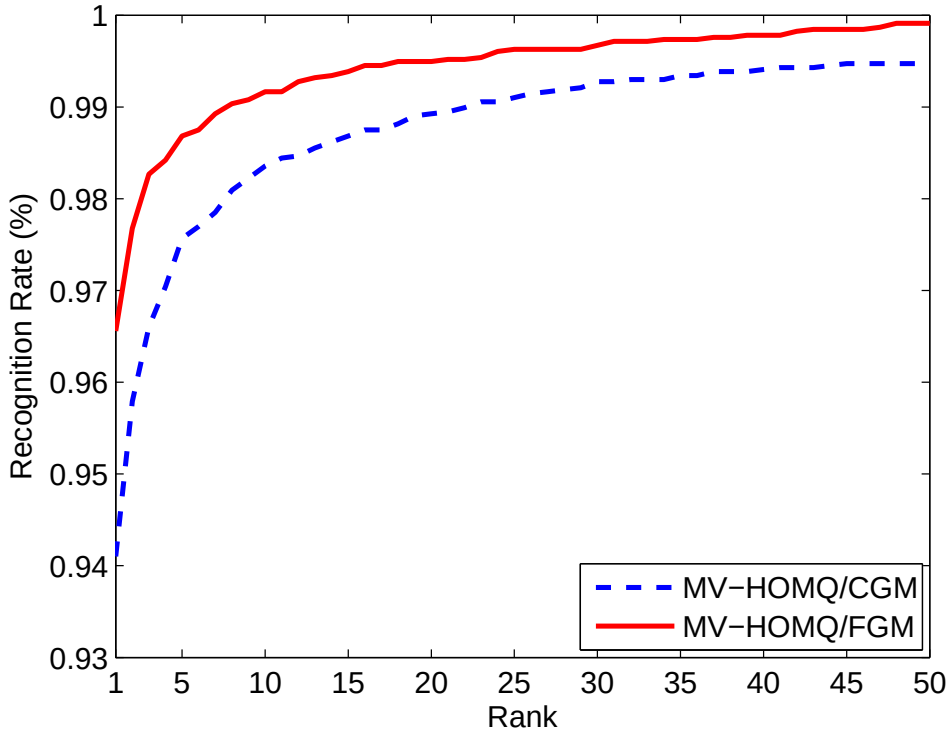


Figure 4.19: *Illustration the Cumulative Match Curves on the Bosphorus database.*

Table 4.11: *Performance comparison on the whole Bosphrous database.*

%	Maes et al. [Maes *et al.* 2010]	MV-HOMQ/FGM
Rank-one score	93.66	**96.56**

4.6.5 Face Recognition on the FRGC v2.0 Database

FRGC v2.0 was designed to evaluate 3D FR algorithms which are widely expected to provide the ability to cope with the unsolved issues, *e.g.*, lighting variations and pose changes, in 2D FR under uncontrolled conditions. However, FRGC v2.0 only contains expressive frontal 3D face scans captured under controlled lighting conditions [Phillips *et al.* 2005]. While the proposed method targets 3D FR under uncontrolled environment, thus making it hard to apply standard preprocessing techniques, *e.g.*, fine 3D face alignment, which are generally adopted by 3D FR techniques evaluated on FRGC v2.0, we also benchmarked the proposed method (MV-HOMQ/FGM) over FRGC v2.0 to check its generalization skill as a recent study shows that 3D FR algorithms with state of the art performance over one database can vastly degrade in another setting using a different dataset [Vijayan *et al.* 2011b]. To test the robustness of our method to expression variations on FRGC v2.0, we adopted the same protocol as in [Mian *et al.* 2008] and divided all the probe set into Neutral and Non-neutral subsets.

4.6.5.1 The Results

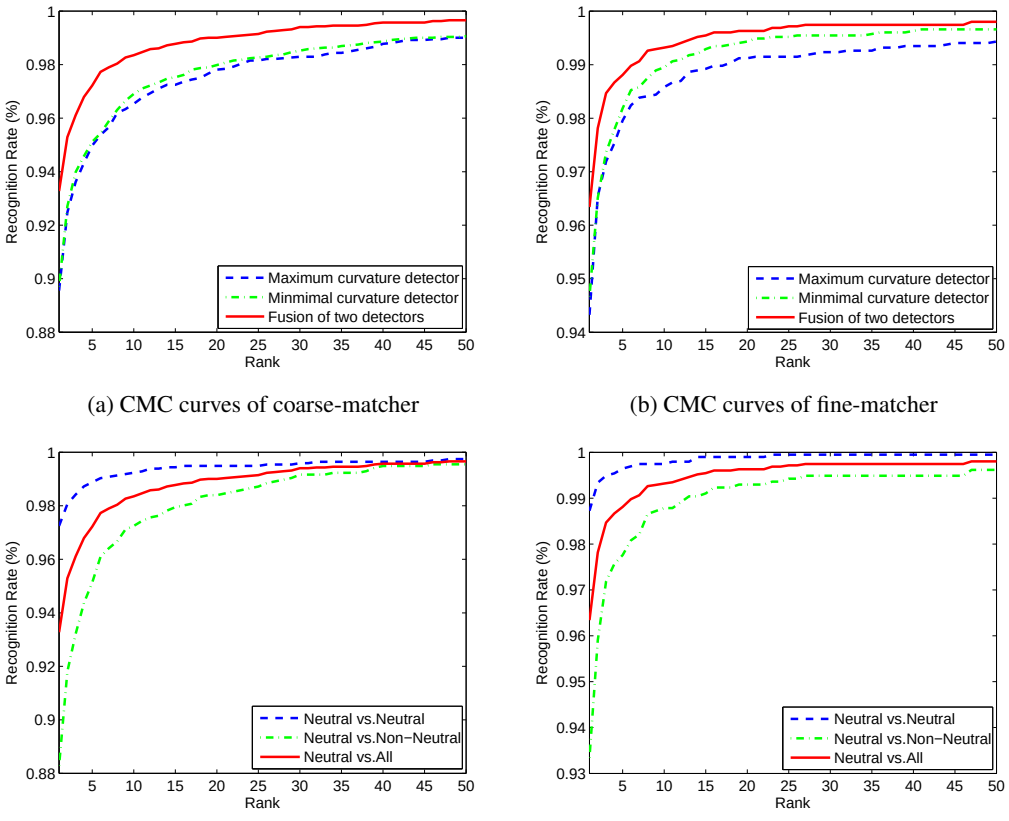
Tab. 4.12 reports the performance of the proposed framework in varying the matcher, *i.e.* CGM vs. FGM, as well as the salient vertex detector, *i.e.* κ_{max} vs. κ_{min} vs. their fusion. As we can see from that table, one can draw the same conclusions as we did on the experimental results on Bosphorus: (a) the proposed method benefits from the joint use of κ_{max} and κ_{min} salient vertex detectors; score-level fusion of two salient detectors for both SIFT-matcher and SRC-matcher is very efficient to improve the recognition rates; (b) The FGM improves a lot the performance of the CGM, especially for the Neutral vs. Non-neutral case; (c) Without any prior registration or subspace-based feature selection (*e.g.*, LDA), the proposed framework MV-HOMQ/SRC achieves competitive results for all the three cases. The corresponding Cumulative Match Curves (CMC) are displayed in Fig. 4.20 and Fig. 4.21.

Table 4.12: *Robustness to Expression Variations on FRGC v2.0.*

%	MV-HOMQ/CGM			MV-HOMQ/FGM		
	κ_{max}	κ_{min}	κ_{max} + κ_{min}	κ_{max}	κ_{min}	κ_{max} + κ_{min}
N vs. N	95.82	95.21	**97.25**	97.91	97.76	**98.73**
N vs. Non-N	81.69	82.78	**88.28**	89.82	90.78	**93.34**
N vs. All	89.56	89.70	**93.28**	94.33	94.67	**96.34**

4.6.5.2 Comparison with the State of the Art

Tab. 4.13 compares the proposed method with several state of the art techniques through the viewpoint of performance drop under facial expression variations whereas Tab. 4.14 lists the state of the art performance in comparison with the proposed method. As we can see from those tables, most state of the art 3D FR algorithms make use of range images and require a more or less costly registration step (*e.g.*, ICP), through several anthropometric landmarks (*e.g.*, nose tip), to align frontal 3D scans with moderate pose variations as those in FRGC v2.0. The best state of the art results so far achieved on FRGC v2.0, *i.e.*, [**?**] [Wang *et al.* 2010] [Spreeuwers 2011], require fine face alignment and/or the learning of discriminating face features through additional 3D face training data. Spreeuwers [Spreeuwers 2011] reports a rank-1 recognition rate as high as 99.0% on the whole FRGC v2.0 database with a method which aligns 3D face scans through an intrinsic face coordinate system thanks to landmark structures, *i.e.* the vertical symmetry plane, the nose tip and the slope of the bridge of the nose, then fuse several region classifiers trained on additional 3D face data, *e.g.*, Bosphorus, using subspace-based methods. Queirolo *et al.* [**?**] reports a 98.4% rank-1 recognition rate on FRGC v2.0 with a method which first locates on a frontal range image six anthropometric landmarks, *e.g.*,the inner right and left eye corners, the right and left nose corners, and the nose tip and base, and then crop four face regions on which a boosted ICP using Simulated Annealing is performed for matching. All the preprocessing steps, *e.g.*, nose tip, symmetry plane, in those methods are simply made unavailable



(a) CMC curves of coarse-matcher

(b) CMC curves of fine-matcher

(c) CMC curves of coarse-matcher

(d) CMC curves of fine-matcher

Figure 4.20: *The Cumulative Match Curves on FRGC v2.0.*

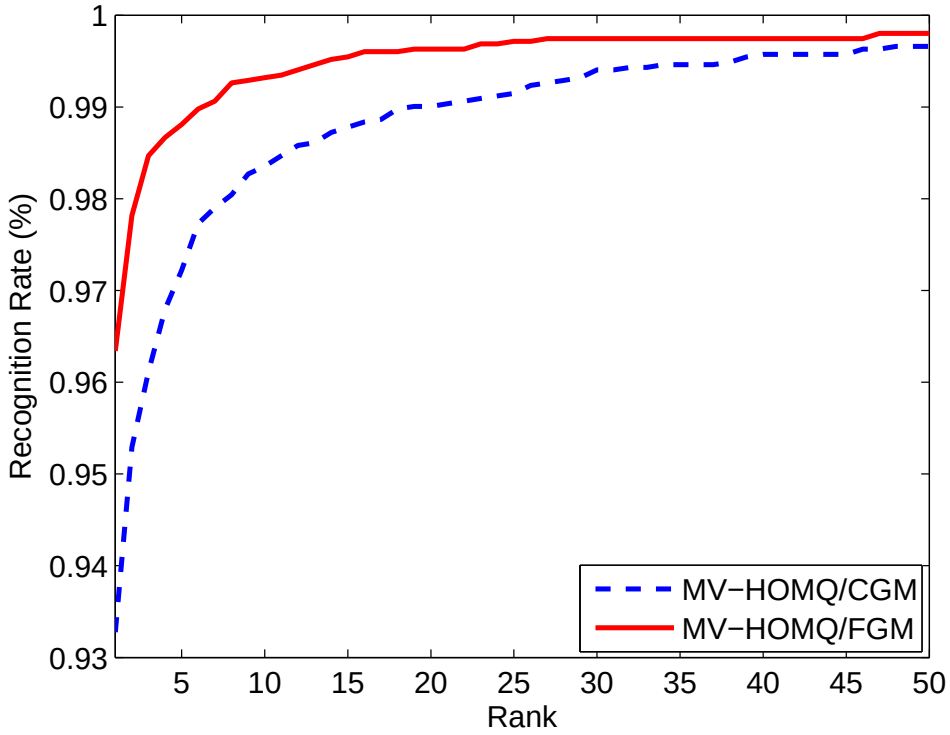


Figure 4.21: *Illustration of the Cumulative Match Curves on the FRGCv2 database.*

because of face missing data due to pose or external occlusions in a scenario of 3D FR under uncontrolled conditions. Once more, without any prior registration and learning for the selection of discriminating features, the proposed framework, MV-HOMQ/FGM as local shape descriptor, achieves a competitive rank-1 recognition rate (96.3%) and consistently demonstrates its effectiveness. Thanks to FGM and the descriptive power of MV-HOMQ, the proposed method also outperforms roughly by 7 points another local feature-based method by Smeets *et al.* [Smeets *et al.* 2013].

4.6.6 Discussion

To gain insights into the proposed local feature-based framework in dealing with those challenges of real-life applications under uncontrolled conditions, we plot in Fig. 4.22 the matching pairs of salient vertices over 3D face scans of the same subject under expression, pose and occlusion variations. In that figure, salient vertices are located by the κ_{max} detector and CGM is used along with MV-HOMQ as local shape descriptor. As we can see from Fig. 4.22, in the case of comparison Neutral vs. Neutral, a large number of salient vertices, spread over the whole facial surface, is accepted as matching ones; When we move to the case of Neutral vs. Expression, all the salient vertices around the opened mouth are automatically rejected in the matching process, while most of the salient vertices detected on the rigid or near rigid regions, *e.g.*, nose, eyes and upper cheek, are retained. For the case of Neutral vs. YR, the matched vertices are concentrated on the visible left part of the facial surfaces when one face scan displays some degree of yaw rotation to the right. Finally,

Table 4.13: *Performance comparison under expression variations on the FRGCv2 database.*

%	N vs. N	N vs. Non-N	Degradation
Mian *et al.* [Mian *et al.* 2008]	99.0	86.7	12.3
Huang *et al.* [Huang *et al.* 2011e]	99.1	92.5	6.6
Huang *et al.* [Huang *et al.* 2011a] [Huang *et al.* 2012]	99.0	94.9	4.1
Li *et al.* [Li *et al.* 2011a]	98.0	94.2	3.8
MV-HOMQ/FGM	98.7	93.3	5.4

Table 4.14: *Performance comparison on the whole FRGC v2.0 database.*

Method	Year	Data format	Face alignment	Rank-one score
Chang *et al.* [Chang *et al.* 2006]	2006	Range	Yes	91.9%
Kakadiaris *et al.* [Kakadiaris *et al.* 2007a]	2007	Range	Yes	91.9%
Mian *et al.* [Mian *et al.* 2007]	2007	Range	Yes	96.2%
Mian *et al.* [Mian *et al.* 2008]	2008	Range	Yes	93.5%
Faltemier *et al.* [Faltemier *et al.* 2008a]	2008	Range	Yes	**97.2%**
Alosaimi *et al.* [Al-Osaimi *et al.* 2009]	2009	Range	Yes	**96.5%**
Alyuz *et al.* [Alyuz *et al.* 2010]	2010	Range	Yes	**97.5%**
Queirolo *et al.* [Queirolo *et al.* 2010]	2010	Range	Yes	**98.4%**
Wang *et al.* [Wang *et al.* 2010]	2010	Range	Yes	**98.3%**
Huang *et al.* [Huang *et al.* 2011a]	2011	Range	No	**97.2%**
Spreeuwers [Spreeuwers 2011]	2011	Range	Yes	**99.0%**
Li *et al.* [Li *et al.* 2011a]	2011	Range	Yes	96.3%
Ballihi *et al.* [Ballihi *et al.* 2012]	2012	Mesh	Yes	98.2%
Smeets *et al.* [Smeets *et al.* 2013]	2013	Mesh	**No**	89.6%
MV-HOMQ/FGM	2013	Mesh	**No**	96.3%

for the case of Neutral vs. Occlusion, there is only a small number of matched vertices around the occluded regions. One can thus easily understand why the performance of CGM which simply counts the number of matching pairs drastically degrades when expression, occlusion, and pose variations occur, since the number of matches is seriously decreased (especially in the case of Neutral vs. YR) compared with that in Neutral vs. Neutral. In contrast, the FGM computes the average reconstruction error between the local shape descriptors of the probe salient vertices and those of gallery ones and largely alleviates the issue of decreasing number of matching pairs in the presence of missing data. Fig. 4.23

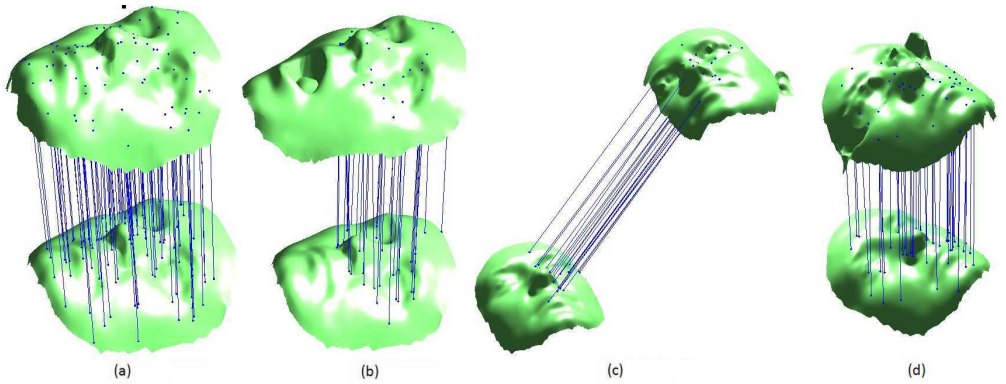


Figure 4.22: *Gaining insights into the proposed local feature-based 3D FR framework. The four cases are displayed using the κ_{max} salient vertex detector and MV-HOMQ/CGM framework: (a) Neutral vs. Neutral; (b) Neutral vs. Expression; (c) Neutral vs. YR; (d) Neutral vs. Occlusion.*

contains the only three 3D face scans in the subset of occlusion which are not correctly recognized by the proposed MV-HOMQ/FGM approach. It can be seen from that figure that hair in these 3D face scans covers up to 50% of the corresponding facial surfaces. Fig. 4.24 gives two other examples of bad recognition in Bosphorus when the extreme pose, *e.g.*, YR90°, leads to more than 50% missing face data.

Figure 4.23: *Examples of non correctly recognized samples with hair occlusion.*

Another problem with the extreme pose is the bad mesh quality of the resulted 3D face scan as shown in Fig. 4.25. As the face is scanned over its right profile, the depth of the scanned facial surface quickly decreases when passing from the closest facial region (ear) to the most remote one (nose) and thereby results in high disparities of triangles of various size. Those disparities make the computation of multiple order differential quantities highly unreliable. Figure 4.26 shows several 3D scans with disgust expressions which are not correctly recognized by the proposed MV-HOMQ/FGM method. It can be seen that the

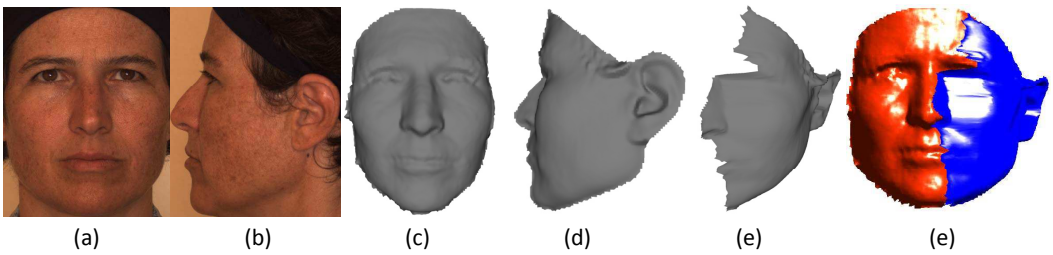


Figure 4.24: *Examples of non recognized faces with large pose variations (YR90°) in Bosphorus.(a): gallery neutral face scan; (b-c):left profile scan (b) and its frontal view (c); (d-e): right profile scan (d) and its frontal view; (e): manual registration of (d) over (a) showing that more than 50% face data is missing in the right profile scan (d).*

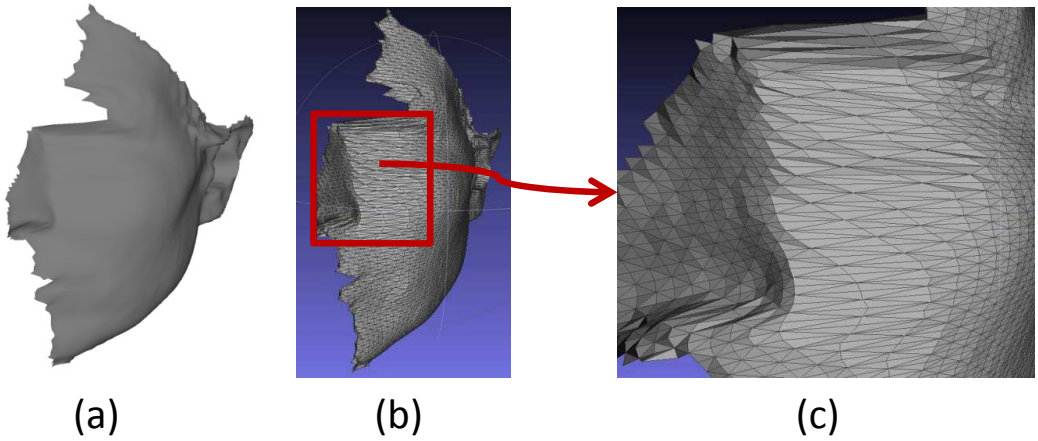


Figure 4.25: *Illustration of the mesh quality of the sample with large pose variation: (a) the scan of BS001YRR90; (b) corresponding triangulation mesh of (a); (c) the zooming of (b) over the region around the eye and nose.*

disgust expressions by those 3D scans are quite exaggerated, leading to serious deformation of the whole facial surfaces, even to the rigid regions e.g. the nose and forehead. While Fig. 4.27 displays some wrong matches caused by exaggerated LAUs, and we can see that similar to the cases in Fig. 4.26, the entire face including the rigid areas are seriously deformed as well. Both the facts indicate when the expression are exaggerated enough to change the relative rigid regions, i.e. the forehead and nose, it will dramatically challenges the proposed method.

Figure 4.28 shows three incorrect matching results by the proposed MV-HOMQ/FGM method. As it can be seen from that figure, exaggerated facial expressions tend to deform the whole facial surfaces and result in very few matching pairs (Fig. 4.28(a)). The same phenomenon occurs on the matching result with a profile 3D face scan (Fig. 4.28(b)). Because of the bad mesh quality which results from the profile 3D face scan, only one (bad) matching pair can be observed on the 3D face scans. In the case of external occlusion (Fig. 4.28(c)), the covering surface induced by the external occlusion (*e.g.*, hair) significantly decrease the number of matching pairs.

To analyze the computational complexity of our approach, we report in Tab. 4.15 the

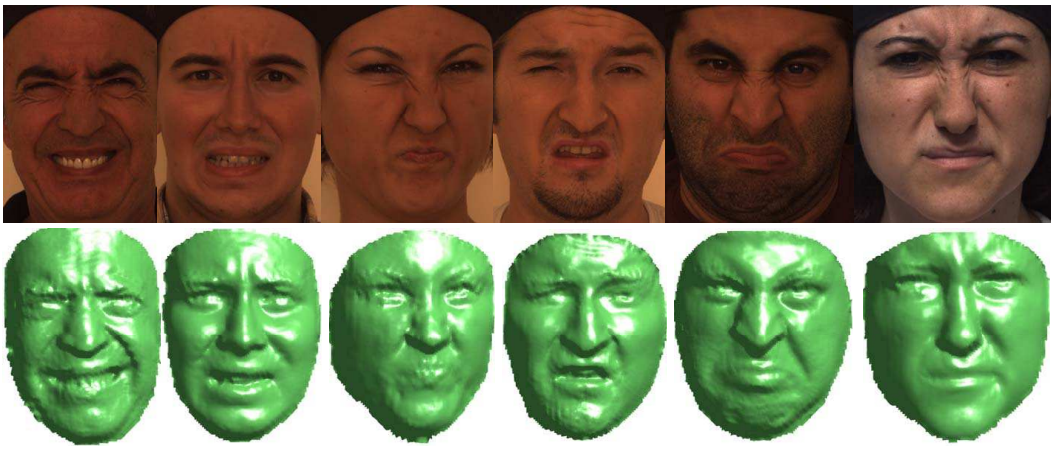


Figure 4.26: *Examples of non correctly recognized samples with disgust expression.*

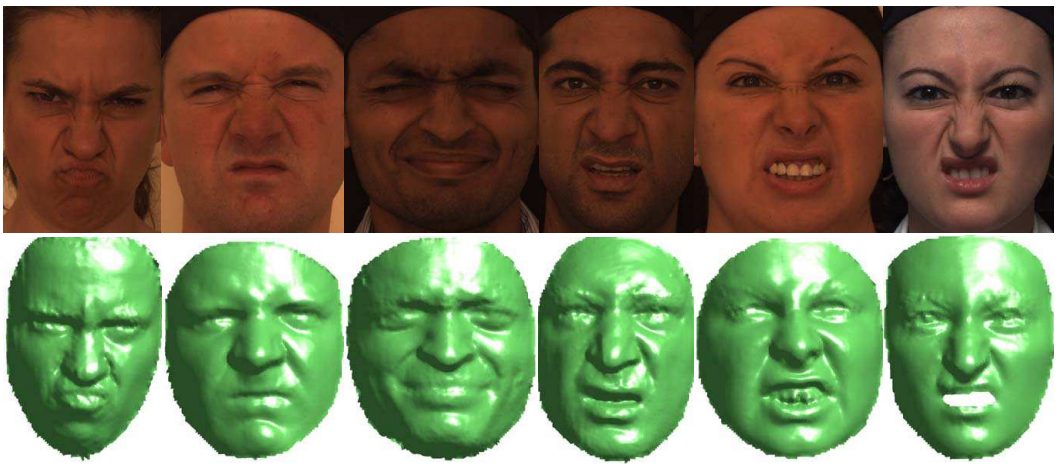


Figure 4.27: *Examples of non correctly recognized samples with lower action unit.*

computation time consumed per salient vertex in each step. Recall that the proposed 3D FR method for the identification of a probe face scan encompasses, as illustrated in Fig. G.1, the following three main stages: the detection of salient vertices, the description of the input face scan as a sparse set of local shape descriptors around the detected salient vertices and the matching of the resulted description with those in the gallery. The proposed 3D FR method was implemented on Matlab[3] on a PC with CPU 950 3.07GHz and 8GB RAM. In particular, the main time cost for salient vertex detection is the estimation of minimum and maximum curvatures over different 3D Gaussian scale spaces of the input mesh. For salient vertex description, the geodesic neighborhood localization step is the most expensive part of the computation. For salient vertex matching, we make use of the notations of section 4.5. Let $\mathbf{D} \in \mathbb{R}^{M \times K}$ be the gallery dictionary. Given a local shape descriptor of a salient vertex $\mathbf{y_i} \in \mathbb{R}^M$ extracted from a probe scan, the theoretical complexity of CGM to identify

[3]Matlab Toolbox Fast Marching is used to compute the geodesic distance and the ring neighborhood vertices.

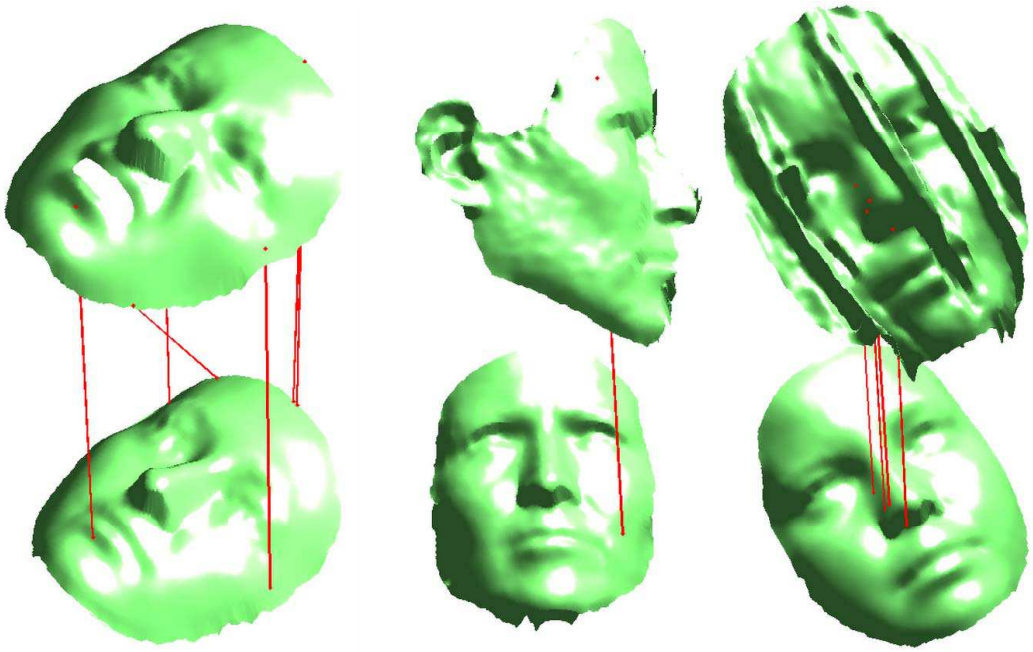


Figure 4.28: *Illustration of three incorrect matching results: (a) Disgust expression; (b) Extreme pose; (c) External occlusion by hair.*

this vertex is to compute K inner products $\langle \mathbf{y_i}, \mathbf{d_j} \rangle$ for all $\mathbf{d_j} \in \mathbf{D}$, which costs $O(M \times K)$; for FGM with $L = 1$, the OMP algorithm also requires to first compute the above K inner products, and then compute the reconstruction error, which costs $O(M)$, thus, totally, FGM costs $O(M \times K) + O(M)$ to match a local shape descriptor of one salient vertex on a probe scan.

As we can see from Tab. 4.15, the most time consuming steps are the detection of salient vertices over a 3D face scan of multiple ten thousands of vertices (*e.g.*, 35K in Bosphorus) and the description of the neighboring regions around those salient vertices, *i.e.*, the computation of the local shape descriptors, respectively. The step of the local shape description deeply depends upon the neighboring distance being used, geodesic or Euclidean distance, with the latter more than ten time faster than the first. It is worth noting that all these computations can be made massively parallel, first because the location of a salient vertex only depends upon its local vicinity in neighboring scales of curvatures, then the extraction of local shape descriptors around salient vertices, as well as the matching of the local shape descriptors of salient vertices with those gallery ones are independent tasks individually, and can be carried out in parallel.

4.7 Conclusion

In this chapter, we introduced a local feature-based 3D FR method to cope with the various challenges of 3D FR in the real under un controlled conditions, *i.e.* facial expression variations, missing face data due to arbitrary pose and external occlusions. In contrast to the overwhelming majority of 3D FR algorithms of the literature which need accurate and cost-

Table 4.15: *Average computation time consumed per salient vertex.*

	Detection	Description		Matching	
		MV-HOMQ		CGM	FGM
		Geodesic	Euclid		
Per Vertex	29ms	170ms	12ms	7ms	13ms

ly registration of face range images, the proposed method is mesh-based and registration free. It describes each 3D face scan as a sparse set of pose invariant local shape descriptors, namely histograms of multiple order surface differential quantities, *i.e.* HOG, HOS, HOGS and their early fusion HOMQ, thereby achieving a compact, comprehensive and pose invariant 3D face description. Those local shape descriptors are computed over the vicinity of salient vertices automatically located by both the κ_{max} and κ_{min} salient vertex detectors in their respective Gaussian scale space. The matching between two 3D face scans is carried out using a SRC-based fine grained matcher (FGM) between their corresponding 3D face descriptions in computing a mean reconstruction error between a probe and a gallery. Without any prior registration nor learning of discriminating face features, the proposed 3D FR framework was comprehensively benchmarked on the Bosphorus database, the largest public 3D face database which best depicts the aforementioned challenges of 3D FR in the real. The experimental results demonstrates the effectiveness of the proposed method which achieves the best state of the art performance in the various scenarios featuring different levels of difficulties with respect to expression variations, pose changes and external occlusions. Additional experiments carried out on the FRGC v2.0 dataset show that the proposed alignment free method consistently displays competitive rank-1 recognition rate and thereby demonstrates its generalization skill over novel face data captured under different conditions.

In the future, we are willing to further reinforce the expressiveness of the proposed pose invariant 3D face description, in considering alternative salient vertex detector, *e.g.*, Harris 3D [Yu *et al.* 2013], and additional local shape descriptors, *e.g.*, spin image, 3D shape context, heat kernel signature [Sun *et al.* 2009]. We also want to study 3D FR using face data of variable quality, *e.g.*, resolution, and improve the proposed 3D FR algorithm over 3D meshes of low quality.

CHAPTER 5

3D Data Quality Improvement: Surface Meshing with Curvature Convergence

Contents

5.1 Introduction . . . 143
5.1.1 Existing Theoretical Results . . . 144
5.1.2 Our Theoretical Results . . . 145
5.2 Previous Works . . . 147
5.2.1 Surface Meshing/Remeshing . . . 147
5.2.2 Conformal Surface Parameterization . . . 148
5.3 Statement of the Main Theorem . . . 148
5.3.1 Curvature Measures . . . 148
5.3.2 Main Results . . . 149
5.4 Theoretical Proofs . . . 150
5.4.1 Surface Uniformization . . . 150
5.4.2 Delaunay Refinement . . . 151
5.4.3 Normal Cycle Theory . . . 153
5.4.4 Estimation . . . 156
5.5 Computational Algorithm . . . 164
5.5.1 Algorithm Pipeline . . . 166
5.6 Experimental Results . . . 167
5.6.1 Triangulation Quality . . . 167
5.6.2 Curvature Measure Comparisons . . . 168
5.7 Conclusion . . . 170

5.1 Introduction

Surface meshing and remeshing play fundamental roles in many engineering fields, including computer graphics, geometric modeling, visualization and medical imaging. Typically, surface meshing finds a set of sample points on the surface with a curved triangulation, then approximates each face by an Euclidean triangle in $\mathbb{R}^3$, thereby approximating the underlying smooth surface by a polyhedral triangular surface, which is called triangle mesh.

Many geometric processing tasks are equivalent to solving geometric partial differential equations (PDEs) on surfaces. The following are some direct examples: for shape analysis, heat kernel signature (HKS) [Sun *et al.* 2009] is mostly utilized, which entails solving a heat equation and computing the eigenvalues and eigenfunctions of the Laplace-Beltrami operator on surfaces; for shape registration, a surface harmonic map [Wang *et al.* 2008] is widely used, which essentially means solving elliptic PDEs on surfaces; for surface parameterization, the discrete Ricci flow [Jin *et al.* 2008] is often computed, which amounts to solving a non-linear parabolic equation on surfaces.

Most geometric PDEs are discretized on triangle meshes, and solved using numerical methods, such as Finite Element Methods (FEM). The numerical stability, the convergence rates, and the approximation bounds of the discrete solutions are largely determined by the quality of the underlying triangle mesh, which is measured mainly by the size and the shape of triangles on the mesh. Therefore, the generation of high quality meshes has fundamental importance.

Most existing meshing and remeshing algorithms are based on the Delaunay refinement method. There are three main different categories:

1. The sampling is computed in $\mathbb{R}^3$, and triangulated using volumetric Delaunay triangulation, such as [Amenta & Bern 1998] [Amenta *et al.* 2000] [Boissonnat & Oudot 2005] [Cheng *et al.* 2004] [Dey *et al.* 2005] [Dey & Levine 2007].

2. The sampling is directly computed on a the smooth surface, such as geodesic Delaunay refinement [Dai *et al.* 2007], and Chew's second algorithm [Paul Chew 1993].

3. The sampling is computed in the conformal parameter domain, and triangulated using planar Delaunay triangulation, such as [Alliez *et al.* 2003] [Remacle *et al.* 2010] [Marchandise *et al.* 2011] [Marchandise *et al.* 2012] [Alliez *et al.* 2008].

The convergence of curvature measures for the approaches of the first two categories has been thoroughly discussed in [Cohen-Steiner & Morvan 2003a] [Cohen-Steiner & Morvan 2006] [Morvan 2008a] [Morvan & Thibert 2004]. This work focuses on the curvature measure convergence of the third category.

5.1.1 Existing Theoretical Results

Based on the classic results of Federer [Federer 1983] and Fu [Fu 1989], among others, the authors in [Cohen-Steiner & Morvan 2003a] [Cohen-Steiner & Morvan 2006] [Morvan 2008a] defined a general and unified framework of curvature measures for both smooth and discrete submanifolds of $\mathbb{R}^N$ based on normal cycle theory. Furthermore, they proved the convergence and approximation theorems of curvature measures for general geometric subset of $\mathbb{R}^N$.

In particular, suppose M is a smooth surface embedded in $\mathbb{R}^3$, M_ε is an ε-*sample* of M, namely, for each point $p \in M$, the ball $B(p, \varepsilon \mathrm{lfs}(p))$ contains at least one sample point in M_ε, where $\mathrm{lfs}(p)$ denotes the local feature size of M at point p. Let T be the triangle mesh induced by the volumetric Delaunay triangulation of M_ε restricted to M. If ε is small enough, each point of the mesh has a unique closest point on the smooth surface. This leads

to the introduction of the closest point projection $\pi : T \to M$. This map has the following properties:

1. Normal deviation: $\forall p \in T,\ |\mathbf{n}(p) - \mathbf{n} \circ \pi(p)| = O(\varepsilon)$, by Amenta *et al.* [Amenta & Bern 1998], and Boissonnat *et al.* [Boissonnat & Oudot 2005].

2. Hausdorff distance: $|p - \pi(p)| = O(\varepsilon^2)$, by Boissonnat *et al.* [Boissonnat & Oudot 2005].

3. Homeomorphism: π is a global homeomorphism, by Amenta *et al.* [Amenta & Bern 1998] and Boissonnat *et al.* [Boissonnat & Oudot 2005].

4. Curvature measures: Let B be a Borel subset of $\mathbb{R}^3$, then the differences between the curvature measures on M and that on T are $K\varepsilon$, where K depends on the triangulation T [Cohen-Steiner & Morvan 2003a], [Morvan 2008a].

In the first category, the authors show that, unfortunately, the convergence of curvature measures can not be guaranteed. Depending on the triangulation, when ε goes to 0, K may go to infinity, (see [Cohen-Steiner & Morvan 2006] for a counterexample). To ensure the convergence of the curvature measures, in [Cohen-Steiner & Morvan 2006] [Morvan 2008a], the authors suggest adding a stronger assumption to the sampling condition, namely, *κ-light ε-sample*, which is an *ε-sample* with the additional constraint that each ball $B(p, \varepsilon \mathrm{lfs}(p))$ contains at most κ sample points.

In the second category, the curvature measure convergence for meshes obtained by Chew's second algorithm on surface has been proved in [Morvan & Thibert 2004]. The normal and area convergence for meshes based on geodesic Delaunay refinement has been proved in [Dai *et al.* 2007]. However, the computation of geodesic Delaunay triangulation is prohibitively expensive [Xin *et al.* 2011] in practice.

5.1.2 Our Theoretical Results

This chapter will deal with triangulations of the third category, showing stronger estimates. Using conformal parameterization, we get meshes satisfying as before the first two properties,

1. Normal deviation: $O(\varepsilon)$, Lemma 5.4.2 and Lemma 5.4.3.

2. Hausdorff distance: $O(\varepsilon^2)$, Lemma 5.4.2 and Lemma 5.4.3.

Moreover, we improve the other two properties as follows:

3. Homeomorphism: Besides the closest point projection π, we also define a novel mapping, the natural projection η, induced by the conformal parameterization. Both projections are global homeomorphisms, see section 4.4.

4. Curvature measures: we show the Delaunay refinement method on conformal parameter domain generates κ-light ε-sample, which guarantees the convergence of curvature measures. Moreover, we show the bounds of the curvature measures are $K\varepsilon$, where K is $O(area(B)) + O(length(\partial B))$, and are independent of the triangulations, see Theorem 5.3.2.

In addition, the coding complexity is much lower and computations are less prohibitive than those in the second category.

5.1.2.1 Similarities

Following the work in [Cohen-Steiner & Morvan 2003a], our proof is mainly based on the normal cycle theory. Both methods estimate both Hausdorff distance and the normal deviation at the corresponding points. Then both methods construct a homeomorphism from the triangle mesh to the surface, which induces a homotopy from the normal cycle of the mesh to the normal cycle of the surface. Then, the volume swept by the homotopy and the area of its boundary are estimated. This gives a bound on the difference between the curvature measures.

5.1.2.2 Differences

But our work has sharp distinctions from theirs, in terms of both theoretical and algorithmic aspects:

- In theory, as pointed out previously, without stronger sampling condition, volumetric Delaunay refinement method cannot guarantee the convergence of curvature measures. In contrast, our results can ensure the convergence curvature measures without extra assumptions.
- In theory, the volumetric Delaunay refinement method requires the embedding of the surface. Our method is intrinsic, which only requires Riemannian metric. In many real-life applications, e.g. the general relativity simulation in theoretical physics, the surface metric is given without any embedding space. In such cases, the volumetric Delaunay method is invalid, but our method can still apply.
- In theory, to prove the main theorem, the closest point mapping is constructed in [Cohen-Steiner & Morvan 2003a]. In contrast, we supply two proofs: one is based on the closest point mapping, whereas the other uses a completely different mapping based on conformal parameterization. Conceptually, besides its novelty, the latter is also simpler.
- In practice, the planar Delaunay refinement method is much easier to implement, the data structure for planar triangulation is much simpler than that of tetrahedral mesh, and the planar algorithm is much more efficient.

Remark

Current meshing algorithm aims to achieve a good triangulation, and requires a conformal parameterization, which in turn requires a triangulation. Consequently, this looks like a chicken-and-egg problem.

In fact, conformal parameterization can be carried out using an initial triangulation of low quality, and this algorithm will produce a new triangulation with much better quality. Many geometric processing tasks cannot be computed on the initial mesh. For example, the error bound for a discrete solution to the Poisson equation is $O(\varepsilon^2)$ on good quality

meshes. If the mesh has too many obtuse angles, then the discrete results will not converge to the smooth solution.

In reality, surfaces are acquired by 3D scanning technologies, such as laser scanner and structured-light scanner, or reconstructed form medical data. Usually, the raw point cloud is very dense, and the initial triangulation is induced by the pixel or voxel grid structures. In the geometric modeling field, the input surfaces may be spline surfaces, and the initial triangulation can be chosen as the regular grids on the parameter domain. Then conformal parameterizations can be computed using the dense samples with the initial triangulation. Then we can perform the remeshing using the current conformal parametric Delaunay refinement algorithm to improve the mesh quality or compress the geometric data.

5.2 Previous Works

5.2.1 Surface Meshing/Remeshing

5.2.1.1 Delaunay Refinement

The Delaunay refinement algorithm was originally designed for meshing planar domains, and was later generalized for meshing surfaces and volumes. We refer readers to [Cheng *et al.* 2012] for full details.

Chew's first algorithm [Paul Chew 1989] splits any triangle whose circumradius is greater than the prescribed shortest edge length parameter ε and hence generates triangulation of uniform density and with no angle smaller than 30°. But the number of triangles produced is not optimal. Chew's second algorithm [Paul Chew 1993] splits any triangle whose circumradius-to-shortest-edge ratio is greater than one, and hence in practice produces grade mesh. Similar split criterion was used in Ruppert's algorithm [Ruppert 1995], which has the theoretical guarantee of the minimal angle of no less than 20.7°. Shewchuk's algorithm [Shewchuk 2001] can create meshes with most angles of 30° or greater. Dey *et al.* developed a series of algorithms for surface meshing and remeshing based on volumetric Delaunay refinement [Cheng *et al.* 2004] [Dey *et al.* 2005] [Dey & Levine 2007], which are the methods in the first category.

5.2.1.2 Centroidal Voronoi Tessellation

The concept of Centroidal Voronoi Tessellations (CVT) was first proposed by Du *et al.* [Du *et al.* 1999], then generalized to constrained Centroidal Voronoi Tessellations [Du *et al.* 2002]. Recently, CVT has been widely used for surface meshing/remeshing to produce high quality triangulations. It can be carried out in the ambient space, e.g. Yan *et al.* [Yan *et al.* 2009], or the conformal parameter domain, e.g. Alliez *et al.* [Alliez *et al.* 2003] [Surazhsky *et al.* 2003], or even high embedding space, e.g. Lévy *et al.* [Lévy & Bonneel 2013]. A complete survey of the recent advancements on CVT based remeshing can be found in [Alliez *et al.* 2008]. Although visually pleasing and uniform, all the existing CVT based remeshing methods for the generation of high quality triangulation have no theoretical bound of the minimal angle [Alliez *et al.* 2008]. Therefore, the curvature measure convergence cannot be guaranteed.

5.2.2 Conformal Surface Parameterization

Over the last two decades, surface parameterization has gradually become a very popular tool for various mesh processing processes [Sheffer *et al.* 2006] [Floater & Hormann 2005]. In this work, we consider only conformal parameterizations. There are many approaches used for this purpose, including harmonic energy minimization [Desbrun *et al.* 2002] [Gu *et al.* 2004] [Wang *et al.* 2005], Cauchy-Riemann equation approximation [Lévy *et al.* 2002], Laplacian operator linearization [Haker *et al.* 2000], circle packing [Hurdal *et al.* 2000], angle-based flattening [Sheffer & de Sturler 2001], holomorphic differentials [Gu & Yau 2003], Ricci curvature flow [Jin *et al.* 2008] [Zeng *et al.* 2010a], Yamabe flow [Lui *et al.* 2010], conformal equivalence class [Springborn *et al.* 2008], most isometric parameterizations (MIPS) [Hormann *et al.* 1999], *etc.*.

5.3 Statement of the Main Theorem

5.3.1 Curvature Measures

First, let M be a C^2-smooth surface embedded in $\mathbb{R}^3$, its curvature measures can be defined as follows.

Définition 2. *The Gaussian curvature measure of M, ϕ_M^G, is the function associated with each Borel set $B \subset \mathbb{R}^3$,*

$$\phi_M^G(B) = \int_{B\cap M} G(p)dp$$

where $G(p)$ is the Gaussian curvature of M at point p. Similarly, the mean curvature measure ϕ_M^H is given by

$$\phi_M^H(B) = \int_{B\cap M} H(p)dp$$

where $H(p)$ denotes the mean curvature of M at point p.

Let V be a polyhedron of $\mathbb{R}^3$ and its polyhedral boundary M be a triangular mesh surface. We use v_i to denote a vertex, $[v_i, v_j]$ an edge, and $[v_i, v_j, v_k]$ a face of M. We define the discrete Gaussian curvature of M at each vertex as the angle deficit,

$$G(v_i) = 2\pi - \sum_{jk} \theta_i^{jk},$$

where θ_i^{jk} is the corner angle on the face $[v_i, v_j, v_k]$ at the vertex v_i. Similarly, the discrete mean curvature at each edge is defined as

$$H(e_{ij}) = |v_i - v_j|\beta(e_{ij}),$$

where β_{ij} is the angle between the normals to the faces incident to e_{ij}. The sign of $\beta(e_{ij})$ is chosen to be positive if e_{ij} is convex and negative if it is concave.

Définition 3. *The discrete Gaussian curvature measure of M, ϕ_M^G, is the function associated with every (Borel) set $B \subset \mathbb{R}^3$*

$$\phi_M^G(B) = \sum_{v\in B\cap M} G(v). \tag{5.1}$$

The discrete mean curvature measure ϕ_M^H *is*

$$\phi_M^H(B) = \sum_{e \in M} length(e \cap B)\beta(e). \tag{5.2}$$

The curvature measures on both smooth surfaces and polyhedral surfaces can be unified by the normal cycle theory, which will be explained in Section 5.4.3.

5.3.2 Main Results

It is well known that any Riemannian metric defined on a smooth (compact with or without boundary) surface M can be conformally deformed into a metric of constant curvature $c \in \{-1, 0, 1\}$, depending on the topology of M, the so-called uniformization metric (cf. Fig. 5.1). Now if M is endowed with a Riemannian metric with constant curvature, Delaunay refinement algorithms can be used to generate a triangulation on M with good quality.

The most common Delaunay refinement algorithms include Chew's [Paul Chew 1989, Paul Chew 1993] and Ruppert's [Ruppert 1995]. Let ε be a user defined upper bound of the circumradius of the final triangulation. Given an initial set of samples on surface M, such that the distance between any pair of samples is greater than ε. If M has boundaries, then the boundaries are sampled and approximated by piecewise geodesics, such that each geodesic segment is greater than ε. The Delaunay refinement method on the uniformization space starts with an initial Delaunay triangulation of the initial samples, then updates the samples by inserting circumcenters of the bad triangles, and meanwhile, updates the triangulation by maintaining the Delaunay property. A bad triangle can be either bad-sized or bad-shaped. A triangle is bad-sized, if its circumradius is greater than ε. A triangle is bad-shaped, if its circumradius-to-shortest-edge ratio is greater than one. In this work, we will show the following meshing algorithm using the packing argument.

Theorem 5.3.1 (Delaunay Refinement)**.** *Let M be a compact Riemannian surface with constant curvature. Suppose that the boundary of M is empty or is a union of geodesic circles. For any given small enough $\varepsilon > 0$, the Delaunay refinement algorithm terminates. Moreover, in the resultant triangulation, all triangles are well-sized and well-shaped, that is*

1. *The circumradius of each triangle is not greater than ε.*
2. *The shortest edge length is greater than ε.*

Suppose M is also embedded in $\mathbb{E}^3$ with the induced Euclidean metric. Then M can also be conformally mapped to a surface with uniformization metric, such that all boundaries (if there are any) are mapped to geodesic circles. By running Delaunay refinement on the uniformization metric, we can get a triangulation of M, which induces a polyhedral surface T, whose vertices are on the surface, and all faces of which are Euclidean triangles. Furthermore, all triangles are well-sized and well-shaped under the original induced Euclidean metric. Based on the induced triangulation T, we will show the following main theorem.

Theorem 5.3.2 (Main Theorem)**.** *Let M be a compact Riemannian surface embedded in $\mathbb{E}^3$ with the induced Euclidean metric, T the triangulation generated by Delaunay refinement*

on conformal uniformization domain, with a small enough circumradius bound ε. If B is the relative interior of a union of triangles of T, then:

$$|\phi_T^G(B) - \phi_M^G(\pi(B))| \leq K\varepsilon \tag{5.3}$$
$$|\phi_T^H(B) - \phi_M^H(\pi(B))| \leq K\varepsilon \tag{5.4}$$
$$|\phi_T^G(B) - \phi_M^G(\eta(B))| \leq K\varepsilon \tag{5.5}$$
$$|\phi_T^H(B) - \phi_M^H(\eta(B))| \leq K\varepsilon \tag{5.6}$$

where for fixed M

$$K = O(\sum_{\{t\in T, t\subset \bar{B}\}} r(t)^2) + O(\sum_{\{t\in T, t\subset \bar{B}, t\cap \partial B\neq \emptyset\}} r(t)),$$

$r(t)$ being the circumradius of triangle t. Moreover, K can be further replaced by: $K = O(area(B)) + O(length(\partial B))$. Furthermore, If M is an abstract compact Riemannian surface (only with a Riemannian metric, but not an embedding), inequality (5.5) still holds.

Here π denotes the closest point projection on M, and η denotes the natural projection on M, which is induced by the conformal parameterization.

5.4 Theoretical Proofs

5.4.1 Surface Uniformization

Let $(M_1, \mathbf{g}_1)$ and $(M_2, \mathbf{g}_2)$ be smooth surfaces with Riemannian metrics. Let $\phi : M_1 \to M_2$ be a diffeomorphism, ϕ is conformal if and only if

$$\phi^* \mathbf{g}_2 = e^{2\lambda} \mathbf{g}_1,$$

where $\phi^* \mathbf{g}_2$ is the pullback metric on M_1, and $\lambda : M_1 \to \mathbb{R}$ is a scalar function defined on M_1. Conformal mappings preserve angles and distort area elements. The *conformal factor* function $e^{2\lambda}$ indicates the area distortion.

According to the classical surface uniformization theorem, every metric surface $(M, \mathbf{g})$ can deform to one of three canonical shapes, the sphere, the Euclidean plane or the hyperbolic plane. Namely, there exists a unique conformal factor function $\lambda : M \to \mathbb{R}$, such that the uniformization Riemannian metric $e^{2\lambda}\mathbf{g}$ induces constant Gaussian curvature, the constant being one of $\{+1, 0, -1\}$ according to the topology of the surface. If surfaces have boundaries, then the boundaries are mapped to circles on the uniformization space.

Figures 5.1 and 5.2 show the uniformizations for closed surfaces and surfaces with boundaries, respectively. The left-hand columns show genus zero surfaces, which can conformally deform to the unit sphere with $+1$ curvatures. The middle columns demonstrate genus one surfaces, whose universal covering space is conformally mapped to the Euclidean plane, and the boundaries become circles. The columns on the right illustrate high genus surfaces, whose universal covering space is flattened to the hyperbolic plane, and whose boundaries are mapped to circles.

Surface uniformization can be carried out using the Ricci flow method [Jin *et al.* 2008]. Then we can compute the triangulation of the surface by performing the Delaunay refinement method on the canonical uniformization domain.

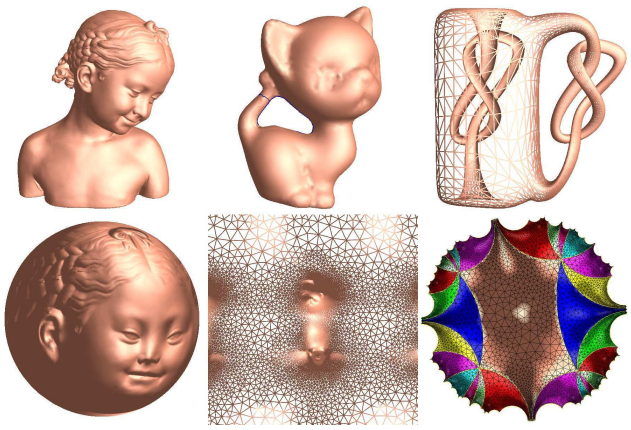


Figure 5.1: *Uniformization for closed surfaces.*

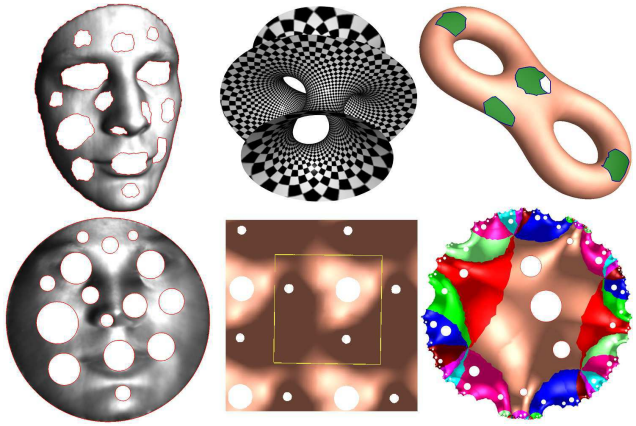


Figure 5.2: *Uniformization for surfaces with boundaries.*

5.4.2 Delaunay Refinement

The Delaunay refinement algorithm for mesh generation operates by maintaining a Delaunay triangulation, which is refined by inserting circumcenters of triangles, until the mesh meets constraints on element quality and size.

5.4.2.1 Geodesic Delaunay Triangulation

By the uniformization theorem, all oriented metric surfaces can be conformally deformed to one of three canonical shapes, the unit sphere $\mathbb{S}^2$, the flat torus $\mathbb{E}^2/\Gamma$ and the hyperbolic surface $\mathbb{H}^2/\Gamma$, where $\mathbb{E}^2$ is the Euclidean plane, $\mathbb{H}^2$ the hyperbolic plane, and Γ is the Deck transformation group, a subgroup of isometries of $\mathbb{E}^2$ or $\mathbb{H}^2$, respectively. The unit sphere $\mathbb{S}^2$ can be conformally mapped to the complex plane by stereographic projection, with a Riemannian metric

$$\mathbb{C}\cup\{\infty\}, \mathbf{g} = \frac{4dzd\bar{z}}{(1+z\bar{z})^2}.$$

Similarly, the hyperbolic plane $\mathbb{H}^2$ is represented by Poincaré's disk model with a Riemannian metric

$$\{|z| < 1 | z \in \mathbb{C}\}, \mathbf{g} = \frac{4dzd\bar{z}}{(1-z\bar{z})^2}.$$

The concepts of Euclidean triangles and Euclidean circles can be generalized to geodesic triangles and geodesic circles on $\mathbb{S}^2$ and $\mathbb{H}^2$. Therefore Delaunay triangulation can be directly defined on these canonical constant curvature surfaces. A triangulation is Delaunay if it satisfies the empty circle property, namely the geodesic circumcircle of each geodesic triangle does not include any other point. Spherical circles on $\mathbb{S}^2$ are mapped to Euclidean circles or straight lines on the plane by stereographic projection. Similarly, hyperbolic circles are mapped to the Euclidean circles on the Poincaré disk. Therefore, geodesic Delaunay triangulations on $\mathbb{S}^2$ or $\mathbb{H}^2$ are mapped to Euclidean Delaunay triangulations on the plane. As a result, geodesic Delaunay triangulations can be carried out using conventional Euclidean Delaunay triangulation.

5.4.2.2 Delaunay Refinement on Constant Curvature Surfaces

The Delaunay refinement algorithm on constant curvature surfaces with empty boundary is introduced as follows. Take a flat torus $\mathbb{E}^2/\Gamma$ as an example. The user chooses a parameter ε, which is the upper bound of the circumradius.

1. An initial set of samples is generated on the surface, such that the shortest distance between any pair of samples is greater than ε. An initial Delaunay triangulation is constructed.

2. Select bad size triangles, whose circumradii are greater than ε, insert their circumcenters, and maintain the Delaunay triangulation.

3. Select bad shape triangles, whose ratio between circum radius and shortest edge length is greater than one, insert their circum centers, maintain the Delaunay triangulation.

4. Repeat 2 and 3, until the algorithm terminates.

The proof of theorem 5.3.1 is based on the conventional packing argument [Paul Chew 1989].

Proof. In the initial setting, all the edge lengths are greater than ε. In step 2, after inserting the circumcenter of a bad size triangle, all the newly generated edges are connected to the center, their lengths are no less than the circumradius, which is greater than ε. In step 3, the circumradius of the bad shape triangle is greater than the shortest edge of the bad triangle, which is greater than ε. All the newly generated edges connecting to the center are longer than the radius ε. Therefore, during the refinement process, the shortest edge is always greater then ε.

Suppose p and q are the closest pair of vertices, then the line segment connecting them must be an edge of the final Delaunay triangulation, which is longer than ε. Therefore, the distance between any pair of vertices is greater than ε. Centered at the each vertex of the triangulation, a disk with radius $\varepsilon/2$ can be drawn. All these disks are disjoint. Because the

total surface area is finite, the number of vertices is finite. Therefore, the whole algorithm will terminate.

When the algorithm terminates, all triangles are well-sized and well-shaped. Namely, the circumradius of each triangle is smaller than ε, and the shortest edge length is greater than ε. For the flat torus case, the minimal angle is greater than 30°. □

By the uniformization theorem, if a surface has a boundary, it can be conformally mapped to the constant curvature surfaces with circular holes. Then the boundaries can be approximated by Planar Straight Line Graph (PSLG), such that the angles between two adjacent segments are greater than 60°. Using a proof similar to the one given by Chew in [Paul Chew 1989] and [Paul Chew 1993], we can show the theorem still holds.

5.4.2.3 Delaunay Refinement on General Surfaces

For general surfaces, we need to add grading to the Delaunay triangulation. The grading function is the conformal factor $e^{2\lambda}$, which controls the size of the triangles. Step 2 in the above algorithm needs to be modified as follows: select a bad size triangle with the circumcenter p and circumradius greater than $\varepsilon e^{-\lambda(p)}$. The same proof can be applied to show the termination of the algorithm. In the resultant triangulation, the grading is controlled by the conformal factor, the circum-radius is less than $\varepsilon e^{-\lambda}$, the shortest edge is greater than $\varepsilon e^{-\lambda}$, so the triangles are still well-shaped. On the original surface, the edge length is greater than ε and the circumradius is less than ε. The minimal angle is bounded.

According to [Funke & Ramos 2002], such a kind of sampling is locally uniform, thus is also a κ-*light* ε-*sample*. Suppose the triangulation is T, $t \in T$ is a triangle, with circum-radius $r(t)$, $B \subset T$ is a union of triangles of T, then

Lemma 5.4.1. *The following estimation holds*

$$\sum_{t \subset \bar{B}} r(t)^2 + \sum_{t \subset \bar{B}, t \cap \partial B \neq \emptyset} r(t) = O(area(B)) + O(length(\partial B)). \tag{5.7}$$

In the following, we denote V as a compact set in three dimensional Euclidean space $\mathbb{R}^3$, M its boundary surface. Hence, M is orientable.

5.4.3 Normal Cycle Theory

In order to be complete, we briefly introduce the normal cycle theory, which closely follows the work in [Cohen-Steiner & Morvan 2003a]. For a more in-depth treatment, we refer readers to [Cohen-Steiner & Morvan 2003a].

Intuitively, the normal cycle of a surface is its offset surface embedded in a higher dimensional Euclidean space. If the surface is not convex nor smooth, its offset surface in $\mathbb{R}^3$ may have self-intersections. By embedding it in higher dimensional space, it can be fully unwrapped.

5.4.3.1 Offset Surface

Suppose V is a volumetric domain in $\mathbb{R}^3$, whose boundary $M = \partial V$ is a compact C^2-smooth surface. Let ρ be the distance between M and the medial axis of the complement of V.

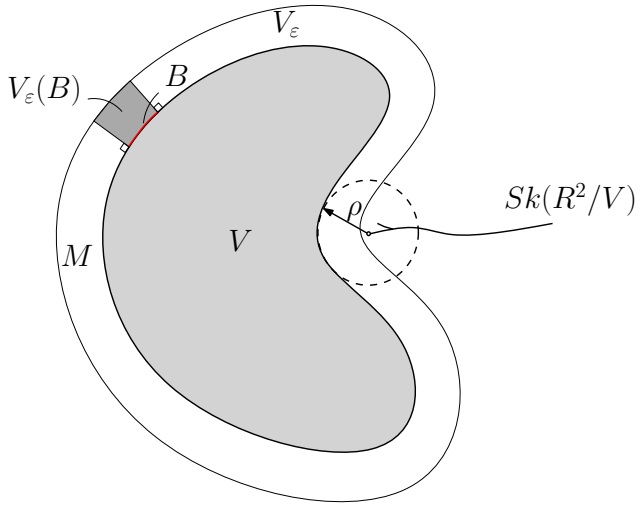


Figure 5.3: *Offset surface and tube formula.*

The ε-offset of V minus V is

$$V_\varepsilon = \{p | p \notin V \; d(p,V) < \varepsilon\} \subset \mathbb{R}^3.$$

The tube formula can be written as

$$Vol(V_\varepsilon) = area(M)\varepsilon + \phi_V^H(M)\frac{\varepsilon^2}{2} + \phi_V^G(M)\frac{\varepsilon^3}{3}$$

for $\varepsilon < \rho$. The localized version of the tube formula is as follows. Let $B \subset M$ be a Borel set, the ε-offset of B is $V_\varepsilon(B)$, then we have

$$Vol(V_\varepsilon(B)) = area(B)\varepsilon + \phi_V^H(B)\frac{\varepsilon^2}{2} + \phi_V^G(B)\frac{\varepsilon^3}{3}.$$

The volume of the ε-offset $V_\varepsilon(B)$ is a polynomial in ε, and its coefficients are multiples of the curvature measures of B. Even if the boundary of V is not smooth but if $\rho > 0$, the volume of $V_\varepsilon(B)$ is always a polynomial in ε for $\varepsilon < \rho$. Therefore the coefficients of this polynomial generalize the curvature measures from smooth surfaces to polyhedral surfaces.

This approach does not generalize to non-convex polyhedral surfaces, where ρ may be equal to 0. So the normal cycle theory has been developed. Intuitively, normal cycles provide a way of unfolding offsets in a higher dimensional space.

5.4.3.2 Normal Cycles

Définition 4. *The normal cycle $N(M)$ of a C^2-smooth surface M is the current associated with the set*

$$N(M) := \{(p, \mathbf{n}(p)) | p \in M\}$$

endowed with the orientation induced by that of M, where a current is the generalization of an oriented surface patch, with integral coefficients. When no confusion is possible, we use the same notation $N(M)$ to denote both the current and its associated set.

The normal cycle of V is the same as that of M, namely, $N(V) = N(M)$. The diffeomorphic mapping from M to its normal cycle $N(M)$ is denoted as

$$\begin{array}{rcl} i & : & M \to N(M) \\ & & p \to (p, \mathbf{n}(p)) \end{array}$$

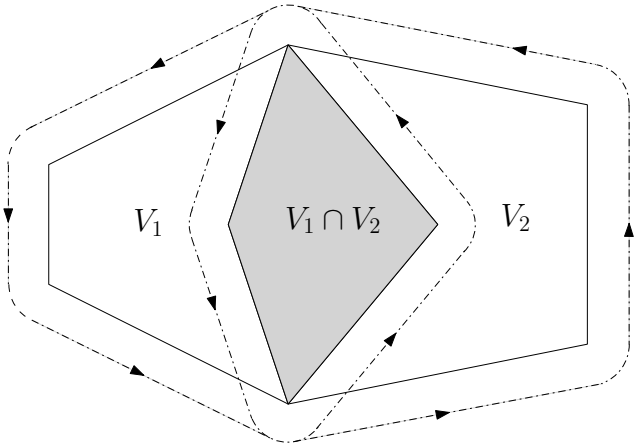


Figure 5.4: *Additivity of the normal cycle.*

Suppose V is a convex body, whose boundary M is a polyhedral surface. We use normal cones to replace normal vectors.

Définition 5. *The normal cone $NC_V(p)$ of a point $p \in V$ is the set of unit vectors v such that*

$$\forall q \in V, \langle q - p, v \rangle \leq 0.$$

Définition 6. *The normal cycle of M is the current associated with the set*

$$\{(p, \mathbf{n}(p)) | p \in M, \mathbf{n} \in NC_V(p)\}$$

endowed with the orientation induced by the one of M.

As in figure 5.4, normal cycles are graphically represented by their image under the map sending $(p, \mathbf{n}(p))$ to $p + \mathbf{n}(p)$.

The crucial property of the normal cycle is its additivity as shown in Fig. 5.4. Suppose V_1 and V_2 are two convex bodies in $\mathbb{R}^3$, such that $V_1 \cup V_2$ is convex, then

$$N(V_1 \cap V_2) + N(V_1 \cup V_2) = N(V_1) + N(V_2).$$

By the additivity property, we can define the normal cycle of a polyhedron. Given a triangulation of the polyhedron V into tetrahedra t_i. $i = 1, 2, \cdots, n$, the normal cycle of V is defined as

$$N(V) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} N(\cap_{j=1}^{k} t_{i_j})$$

by inclusion-exclusion. It is proved that the normal cycle $N(V)$ is independent of triangulations.

Similar to the smooth surface case, one can define a set-valued mapping from M and its normal cycle $N(M)$

$$\begin{array}{ll} i & : M \to N(M) \\ & p \to (p, \mathbf{n}(p)) \quad \mathbf{n} \in NC_V(p). \end{array}$$

5.4.3.3 Invariant Differential 2-Forms

Normal cycles are embedded in the space $\mathbb{R}^3 \times \mathbb{R}^3$, denoted as $E_p \times E_n$, where E_p is called *point space*, and E_n is called *normal space*. Let g be a rigid motion of $\mathbb{R}^3$, $g(p) = Rp + d$, where R is a rotation matrix, d is a translation vector. g can be extended to $E_p \times E_n$ as $\hat{g}(p,n) = (R(p) + d, R(n))$. We say that a differential 2-form ω is invariant under rigid motions, if

$$\hat{g}^* \omega = \omega.$$

The following invariant 2-forms play fundamental roles in normal cycle theory,

Définition 7. *Let the coordinates of $E_p \times E_n$ be $(x^1, x^2, x^3, y^1, y^2, y^3)$, then*

$$\begin{array}{rcl} \omega^A & = & y^1 dx^2 \wedge dx^3 + y^2 dx^3 \wedge dx^1 + y^3 dx^1 \wedge dx^2 \\ \omega^G & = & y^1 dy^2 \wedge dy^3 + y^2 dy^3 \wedge dy^1 + y^3 dy^1 \wedge dy^2 \\ \omega^H & = & y^1 (dx^2 \wedge dy^3 + dy^2 \wedge dx^3) + \\ & & y^2 (dx^3 \wedge dy^1 + dy^3 \wedge dx^1) + \\ & & y^3 (dx^1 \wedge dy^2 + dy^1 \wedge dx^2). \end{array}$$

Curvature measures of a surface can be recovered by integrating specific differential forms on its normal cycle. The following formula unifies the curvature measures on both smooth surfaces and polyhedral surfaces. For a Borel set $B \subset \mathbb{R}^3$, the curvature measures are given by

$$\begin{array}{rcl} \int_{N(M)} \omega^G_{|i(B \cap M)} & = & \phi^G_M(B) \\ \int_{N(M)} \omega^H_{|i(B \cap M)} & = & \phi^H_M(B) \\ \int_{N(M)} \omega^A_{|i(B \cap M)} & = & area(B) \end{array}$$

where $\omega^G_{|i(B \cap M)}$ denotes the restriction of ω to $i(B \cap M)$.

5.4.4 Estimation

In this section, we explicitly estimate the Hausdorff distance, normal deviation, and differences of curvature measures between the discrete triangular mesh to the smooth surface.

5.4.4.1 Configuration

Let $(M, \mathbf{g})$ be a C^2 metric surface. $\mathscr{D}$ is the unit disk on the uv-plane. A conformal parameterization is given by $\varphi : \mathscr{D} \to M$, such that $\mathbf{g}(u,v) = e^{2\lambda(u,v)}(du^2 + dv^2)$. Suppose $p \in \mathscr{D}$ is a point on the parameter domain, then $\varphi(p)$ is a point on the surface. The derivative map $d\varphi|_p : T_p\mathscr{D} \to T_{\varphi(p)}M$ is a linear map

$$d\varphi|_p = e^{\lambda(p)} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

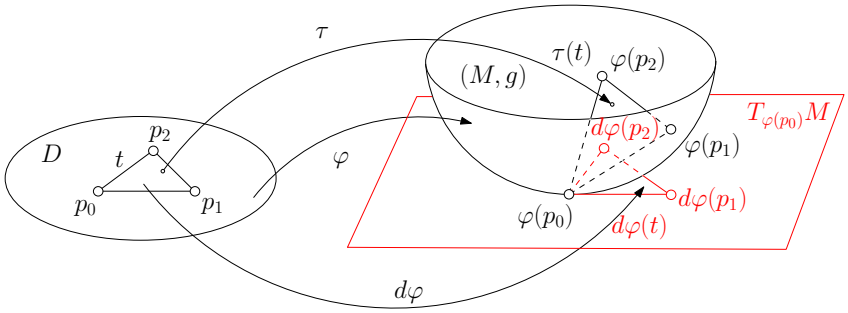


Figure 5.5: *Configuration.*

Let $\mathscr{T}$ be a triangulation of the parameter domain $\mathscr{D}$, the conformal mapping φ induces a triangulation $\varphi(\mathscr{T})$ of the smooth surface M, T is the corresponding polyhedral surface. Each vertex $p_i \in \mathscr{T}$ corresponds to a vertex $\varphi(p_i) \in T$, each triangle on $\mathscr{T}$ corresponds to a face in T. This induces a piecewise linear map $\tau : \mathscr{T} \to T$ (see Fig. 5.5 and Tab. 5.1).

Définition 8 (Natural projection)**.** *The homeomorphism from the triangle mesh T to the smooth surface M*

$$\eta = \varphi \circ \tau^{-1} : T \to M$$

is called the natural projection.

Another map from the mesh to the surface is the closest point projection,

Définition 9 (Closest point projection)**.** *Suppose T has no intersection with the medical axis of M. Let $q \in T$, and $\pi(q)$ be its closest point on the surface M,*

$$\pi(q) = argmin_{r \in M} |r - q|,$$

we call the mapping from q to its closest point $\pi(q)$ as the closest point projection*.*

We will show that the closest point projection is also a homeomorphism.

5.4.4.2 Hausdorff Distance and Normal Deviation

In the following discussion, we assume the triangulation is generated by the Delaunay Refinement in Theorem 5.3.1. Our goal is to estimate the Hausdorff distance and the normal deviation, in terms of both natural projection and closest point projection.

Lemma 5.4.2 (Natural projection)**.** *Suppose $q \in T$, then*

$$|q - \eta(q)| = O(\varepsilon^2), \tag{5.8}$$

$$|\mathbf{n}(q) - \mathbf{n}(\eta(q))| = O(\varepsilon). \tag{5.9}$$

Symbols	*Meaning*
Surfaces	
M	smooth metric surface
$\mathscr{D}$	parameter domain
Mappings	
$\varphi : \mathscr{D} \to M$	conformal parameterization
$d\varphi\|_p : T_p\mathscr{D} \to T_{\varphi(p)}M$	derivative map of φ
$\tau : \mathscr{T} \to T$	piecewise linear map between meshes
$\eta = \varphi \circ \tau^{-1} : T \to M$	natural projection from T and M
$\pi : T \to M$	closest point projection from T and M
Triangulations	
$\mathscr{T}$	triangulation on $\mathscr{D}$
$\varphi(\mathscr{T})$	surface triangulation induced by φ
T	polyhedral surface induced by $\varphi(\mathscr{T})$
Faces	
$t \subset \mathscr{D}$	a triangle on parameter domain
$d\varphi(t) \subset T_{\varphi(p)}M$	corresponding face on tangent plane
$\tau(t) \subset T$	corresponding face on T

Table 5.1: *Symbol list*

Proof. As shown in Fig. 5.5, suppose $p \in \mathscr{D}$, $\tau(p) = q$. p is inside a triangle $t = [p_0, p_1, p_2]$,

$$p = \sum_{k=0}^{2} \alpha_k p_k, 0 \leq \alpha_k \leq 1,$$

where α_k's are barycentric coordinates. All the edge lengths are $\Theta(\varepsilon)$, and angles are bounded. The area is $\Theta(\varepsilon^2)$.
Equation 5.8: By the linearity of τ and $d\varphi$, $\tau(p_k) = \varphi(p_k)$ and $|\varphi(p_k) - d\varphi(p_k)| = O(\varepsilon^2)$, we obtain

$$\begin{aligned} |\tau(p) - d\varphi(p)| &= |\textstyle\sum_k \alpha_k(\tau(p_k) - d\varphi(p_k))| \\ &\leq \textstyle\sum_k \alpha_k |\varphi(p_k) - d\varphi(p_k)| \\ &= O(\varepsilon^2). \end{aligned}$$

Therefore

$$|\tau(p) - \varphi(p)| \leq |\tau(p) - d\varphi(p)| + |d\varphi(p) - \varphi(p)| = O(\varepsilon^2),$$

where $q = \tau(p)$ and $\eta(q) = \varphi \circ \tau^{-1}(q) = \varphi(p)$, this gives Eqn. 5.8.
Equation 5.9: Construct local coordinates on the tangent plane $T_{\varphi(p_0)}M$, such that $\varphi(p_0)$ is at the origin, $d\varphi(p_1)$ is along the x-axis. Then $\tau(p_1)$ is $(\Theta(\varepsilon), 0, O(\varepsilon^2))$, $\tau(p_2)$ is $(\Theta(\varepsilon)\cos\beta, \Theta(\varepsilon)\sin\beta, O(\varepsilon^2))$, where β is the angle at p_0. By direct computation, the normal to the face $\tau(t)$ is $(O(\varepsilon), O(\varepsilon), \Theta(1))$. Therefore

$$|\mathbf{n} \circ \tau(p) - \mathbf{n} \circ \varphi(p_0)| = O(\varepsilon).$$

Furthermore,

$$\begin{aligned} |\mathbf{n} \circ \varphi(p) - \mathbf{n} \circ \varphi(p_0)| &= |W(\varphi(p) - \varphi(p_0))| \\ &\leq \|W\| |\varphi(p) - \varphi(p_0)| \\ &= O(\varepsilon), \end{aligned}$$

where W is the Weigarten map. M is compact, therefore $\|W\|$ is bounded, $|\varphi(p)-\varphi(p_0)|$ is $O(\varepsilon)$.

$$\begin{aligned} |\mathbf{n}\circ\tau(p)-\mathbf{n}\circ\varphi(p)| &\leq |\mathbf{n}\circ\varphi(p)-\mathbf{n}\circ\varphi(p_0)| \\ &+ |\mathbf{n}\circ\tau(p)-\mathbf{n}\circ\varphi(p_0)| \\ &= O(\varepsilon). \end{aligned}$$

This gives Eqn. 5.9. □

Lemma 5.4.3 (Closest point projection)**.** *Suppose $q \in T$, then*

$$|q-\pi(q)| = O(\varepsilon^2), \tag{5.10}$$

$$|\mathbf{n}(q)-\mathbf{n}(\pi(q))| = O(\varepsilon). \tag{5.11}$$

Proof. Equation 5.10: From Eqn. 5.8 and the definition of closest point, we obtain

$$|q-\pi(q)| \leq |q-\eta(q)| = O(\varepsilon^2).$$

Equation 5.11: From Eqn. 5.8 and Eqn. 5.10, we get

$$|\eta(q)-\pi(q)| \leq |\eta(q)-q| + |q-\pi(q)| = O(\varepsilon^2),$$

therefore

$$|\mathbf{n}\circ\eta(q)-\mathbf{n}\circ\pi(q)| \leq \|W\||\eta(q)-\pi(q)| = O(\varepsilon^2).$$

Then from Eqn. 5.9 and the above equation,

$$\begin{aligned} |\mathbf{n}(q)-\mathbf{n}(\pi(q))| &\leq |\mathbf{n}(q)-\mathbf{n}\circ\eta(q)| \\ &+ |\mathbf{n}\circ\eta(q)-\mathbf{n}\circ\pi(q)| \\ &= O(\varepsilon)+O(\varepsilon^2). \end{aligned}$$

□

Remark The proofs for the Hausdorff distances in Eqn. 5.8 and Eqn. 5.10 do not require the triangulation to be well-shaped, but only well-sized. The proofs for the normal deviation estimation in Eqn. 5.9 and Eqn. 5.11 require the triangulation to be both well-sized and well-shaped. In the proofs we use the facts that the triangulation on parameter domain has bounded angles, and the mapping φ is conformal. Figure 5.6 shows a counterexample: a triangle is inscribed in a latitudinal circle of a cylinder, no matter how small it is, its normal is always orthogonal to the surface normals.

5.4.4.3 Global Homeomorphism

Both the natural projection and the closest point projection are homeomorphisms. While it is trivial for natural projection, in the following we give a detailed proof to show the closest point projection is a piecewise diffeomorphism, and estimate its Jacobian.

Lemma 5.4.4. *The closest point projection $\pi : T \to M$ is a homeomorphism.*

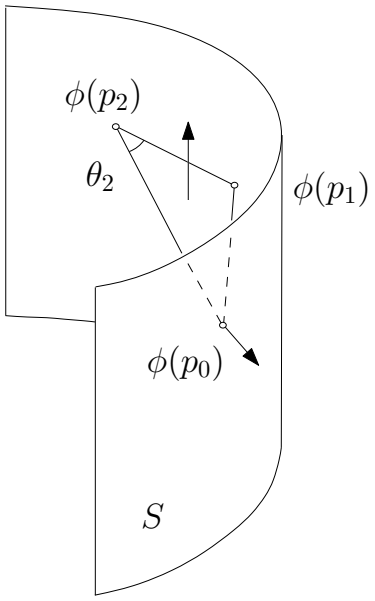


Figure 5.6: *Small triangles inscribed to attitudinal circles of a cylinder do not guarantee the normal convergence.*

Proof. First we show that π restricted to the one-ring neighborhood of each vertex of T is a local homeomorphism. Suppose $p \in T$ is a vertex, therefore $p \in M$ as well. $U(p)$ is the union of all faces adjacent to p. We demonstrate that $\pi : U(p) \to M$ is bijective. Assume $q \in U(p)$, then $|p-q| = O(\varepsilon)$,

$$|\pi(q) - p| \leq |\pi(q) - q| + |q - p| = O(\varepsilon^2) + O(\varepsilon).$$

Therefore

$$|\mathbf{n}(\pi(q)) - \mathbf{n}(p)| = O(\varepsilon). \tag{5.12}$$

Assume there is another point $r \in U(p)$, such that $\pi(q) = \pi(r)$. Let the unit vector of the line segment connecting them be

$$\mathbf{d} = \frac{r-q}{|r-q|},$$

then because $r, q \in U(p)$, $\mathbf{d}$ is almost orthogonal to $\mathbf{n}(p)$,

$$\langle \mathbf{d}, \mathbf{n}(p) \rangle = O(\varepsilon). \tag{5.13}$$

On the other hand, $\mathbf{d}$ is along the normal direction at $\pi(q)$, $\mathbf{n}(\pi(q)) = \pm\mathbf{d}$, assume $\mathbf{d}$ is along $\mathbf{n}(\pi(q))$, from Eqn. 5.12, we obtain

$$|\mathbf{d} - \mathbf{n}(p)| = O(\varepsilon). \tag{5.14}$$

Eqn. 5.13 and Eqn. 5.14 contradict each other. Therefore $\pi_{|U(p)}$ is bijective.

Then we show π restricted on each face is a diffeomorphism. Let $\mathbf{r}(u,v), \mathbf{n}(u,v)$ be position and normals of M respectively, where (u,v) are local parameters along the principal directions. $t \in T$ is a planar face. The inverse closest point projection map is

$\pi^{-1} : \mathbf{r}(u,v) \to \mathbf{q}(u,v)$, where $\mathbf{q}(u,v)$ is the intersection between the ray through $\mathbf{r}(u,v)$ along $\mathbf{n}(u,v)$ and the face t,

$$\mathbf{q}(u,v) = \mathbf{r}(u,v) + s(u,v)\mathbf{n}(u,v),$$

direct computation shows

$$\langle \mathbf{q}_u \times \mathbf{q}_v, \mathbf{n} \rangle = (1 + 2Hs + Ks^2)\langle \mathbf{r}_u \times \mathbf{r}_v, \mathbf{n} \rangle, \tag{5.15}$$

where $s = O(\varepsilon^2)$. When ε is small enough, the above equation is close to 1, which means $\pi_{|U(P)|}$ is a piecewise diffeomorphism.

Secondly, we show that π is a global homeomorphism. We have shown that π is a covering map. At each vertex of T, the closest point equals itself, therefore the degree of π is 1. So π is a global homeomorphism. □

Note that, the estimation of the Jacobian of the closest point projection in Eqn. 5.15 can be applied to show the following. Suppose $B \subset \mathbb{R}^3$ is a Borel set, then

$$|area(B \cap T) - area(\pi(B) \cap M)| = K\varepsilon^2.$$

5.4.4.4 Proof of the Main Theorem

In the following, we give two different proofs of the main theorem 3.4. The first one is based on the following main theorem in [Cohen-Steiner & Morvan 2003a].

Theorem 5.4.5. *Suppose T is a bounded aspect ratio triangulation projecting homeomorphically on M, if B is a relative interior of a union of triangles of T, then*

$$|\phi_T^G(B) - \phi_M^G(\pi(B))| \leq K\varepsilon \tag{5.16}$$
$$|\phi_T^H(B) - \phi_M^H(\pi(B))| \leq K\varepsilon \tag{5.17}$$

where for fixed M

$$K = O(\sum_{\{t \in T, t \subset \bar{B}\}} r(t)^2) + O(\sum_{\{t \in T, t \subset \bar{B}, t \cap \partial B \neq \emptyset\}} r(t)),$$

$r(t)$ is the circumradius of triangle t.

The first proof is a simple corollary of this theorem, based on the Lemma 5.4.4 and the Theorem 5.3.1.

Proof. By Lemma 5.4.4, the closest point projection is a homeomorphism, by Theorem 5.3.1 the triangulation T has bounded aspect ratio, therefore the conditions of Theorem 5.4.5 are satisfied, and consequently, Eqns. 5.16 and 5.17 hold. According to Eqn. 5.7 in Lemma 5.4.1, therefore the main theorem holds. □

The second proof is more direct, and replace the closest point projection $\pi : T \to M$ by the natural projection $\eta : T \to M$.

Proof. The natural projection $\eta : T \to M$ can be lifted to a mapping between the two normal cycles $f : N(T) \to N(M)$, such that the following diagram commutes:

$$\begin{array}{ccc} N(M) & \xleftarrow{f} & N(T) \\ {\scriptstyle i}\uparrow & & \downarrow{\scriptstyle p_1}, \\ M & \xleftarrow{\eta} & T \end{array}$$

where p_1 is the projection from $E_p \times E_n$ to E_p, and $i(q) = (q, \mathbf{n}(q))$ for all $q \in M$. Namely, given a point $q \in T$, and $\mathbf{n}(q)$ in its normal cone, $(q, \mathbf{n}(q)) \in N(T)$,

$$f : (q, \mathbf{n}(q)) \to (\eta(q), \mathbf{n} \circ \eta(q)) \in N(M).$$

By Lemma 5.4.2,

$$|(q, \mathbf{n}(q)) - f(q, \mathbf{n}(q))| = O(\varepsilon). \tag{5.18}$$

It is obvious that f is continuous.

Let $B \subset E_p$, we denote the current $N(T) \cap (B \times E_n)$ by D, and the current $N(M) \cap (\eta(B) \times E_n)$ by E, as shown in Fig. 5.7. Consider the affine homotopy h between f and

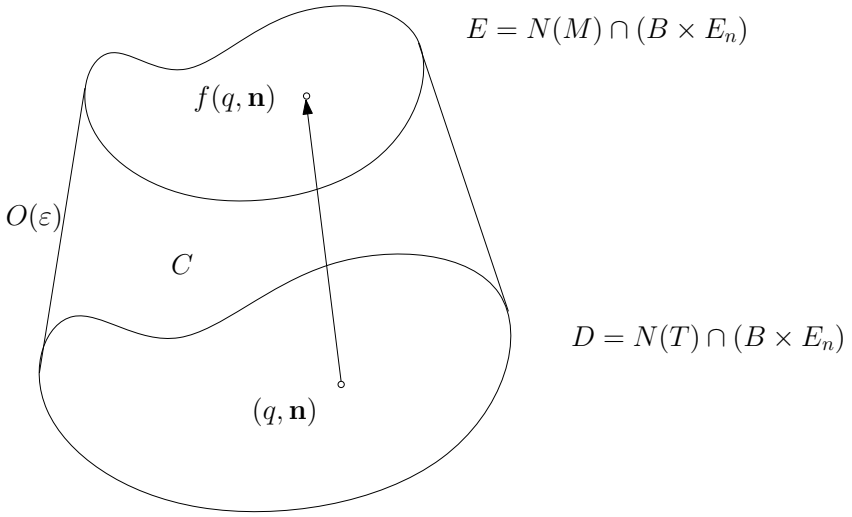


Figure 5.7: *Homotopy between the normal cycles $N(T)$ and $N(M)$.*

the identity,

$$h(t, \cdot) = (1-t)id(\cdot) + tf(\cdot), t \in [0,1].$$

We define the volume swept by the homotopy as

$$C = h_{\#}([0,1] \times D),$$

whose boundary is

$$\partial C = E - D - h_{\#}([0,1] \times \partial D).$$

Intuitively, C is a prism, the ceiling is E, the floor is D, and the walls are $h_{\#}([0,1] \times \partial D)$.

$$\phi_M^G(\eta(B)) - \phi_T^G(B) = \int_{E-D} \omega^G = \int_{\partial C} \omega^G + \int_{h_\#([0,1]\times\partial D)} \omega^G.$$

By Stokes' Theorem,

$$\int_{\partial C} \omega^G = \int_C d\omega^G.$$

Both ω^G and its exterior derivative $d\omega^G$ are bounded, therefore, we need to estimate the volume of block C and the area of the wall $h_\#([0,1]\times\partial D)$. We use $\mathbf{M}(\cdot)$ to denote the flat norm (volume, area, length).

The volume of the prism C is bounded by the height and the section area. The height is bounded by $sup|f-id|$. The section area is bounded by the product of the bottom area $\mathbf{M}(D)$ and the square of the norm

$$\|Dh(t,\cdot)\|^2 = \|tDf + (1-t)id\|^2 \leq (t\sup\|Df\| + (1-t))^2.$$

In later discussion, we will see that $\sup\|Df\| \geq 1$, therefore

$$\|Dh(t,\cdot)\| \leq \sup\|Df\|.$$

We obtain

$$\begin{array}{rcl} \mathbf{M}(C) & \leq & \mathbf{M}(D)sup|f-id|sup\|Df\|^2, \\ \mathbf{M}(h_\#([0,1]\times\partial D)) & \leq & \mathbf{M}(\partial D)sup|f-id|sup\|Df\|. \end{array}$$

Now we estimate each term one by one.

1. Eqn. 5.18 shows
$$sup|f-id| = O(\varepsilon).$$

2. Since the triangulation has a bounded ratio of circumradius to edge length, we obtain
$$\begin{array}{rcl} \mathbf{M}(D) & = & O(\sum_{t\in T, t\subset\bar{B}} r(t)^2) \\ \mathbf{M}(\partial D) & = & O(\sum_{t\in T, t\subset\bar{B}, t\cap\partial B\neq\emptyset} r(t)). \end{array}$$
Let K be the summation of the two terms above. According to Lemma 5.4.1, K is bounded by the area of B and the length of ∂B.

3. For the estimation of $\|Df\|$, we observe that on each triangle $t \in \mathscr{D}$, the mapping τ converges to $d\varphi$, so Df on each triangle converges to
$$(\mathbf{r}_u, 0)du + (\mathbf{r}_v, 0)dv \to (\mathbf{r}_u, \mathbf{n}_u)du + (\mathbf{r}_v, \mathbf{n}_v)dv,$$
where $\mathbf{r}(u,v)$ and $\mathbf{n}(u,v)$ are the position and normal vectors of the smooth surface M, (u,v) the conformal parameters, namely, $|\mathbf{r}_u| = e^\lambda$, $|\mathbf{r}_v| = e^\lambda$ and $\mathbf{r}_u \perp \mathbf{r}_v$.
Assume $(du,dv) = (\cos\theta, \sin\theta)$ for any angle θ, we obtain that the norm of the tangent vector on the left hand side is e^λ. The norm of the vector on the right hand side is bounded by the eigenvalues of the following matrix
$$\left[\begin{array}{cc} \langle(\mathbf{r}_u,\mathbf{n}_u),(\mathbf{r}_u,\mathbf{n}_u)\rangle & \langle(\mathbf{r}_u,\mathbf{n}_u),(\mathbf{r}_v,\mathbf{n}_v)\rangle \\ \langle(\mathbf{r}_v,\mathbf{n}_v),(\mathbf{r}_u,\mathbf{n}_u)\rangle & \langle(\mathbf{r}_v,\mathbf{n}_v),(\mathbf{r}_v,\mathbf{n}_v)\rangle \end{array}\right] = e^{2\lambda} id + III, \tag{5.19}$$

where the third fundamental form is

$$III = \begin{bmatrix} \langle \mathbf{n}_u, \mathbf{n}_u \rangle & \langle \mathbf{n}_u, \mathbf{n}_v \rangle \\ \langle \mathbf{n}_v, \mathbf{n}_u \rangle & \langle \mathbf{n}_v, \mathbf{n}_v \rangle \end{bmatrix}.$$

From $III - 2HII + GI = 0$, where the first fundamental form $I = e^{2\lambda} id$, the second fundamental form $II = e^{2\lambda} W$, W is the Weigarten matrix, we get

$$III = 2HII - GI = e^{2\lambda}(2HW - Gid).$$

Plugging into Eqn. 5.19, we get $\|Df\|^2$ bounded by the eigenvalues of

$$(1-G)id + 2HW,$$

therefore on each face

$$\|Df\|^2 \leq \max\{1+k_1^2, 1+k_2^2\}.$$

So $\|Df\|^2$ is globally bounded.

Putting all the estimates together, we obtain

$$|\phi_M^G(\eta(B)) - \phi_T^G(B)| \leq K\varepsilon.$$

According to Lemma 5.4.1, K is bounded by the area of B and the length of ∂B. The proof for mean curvature measure is exactly the same. □

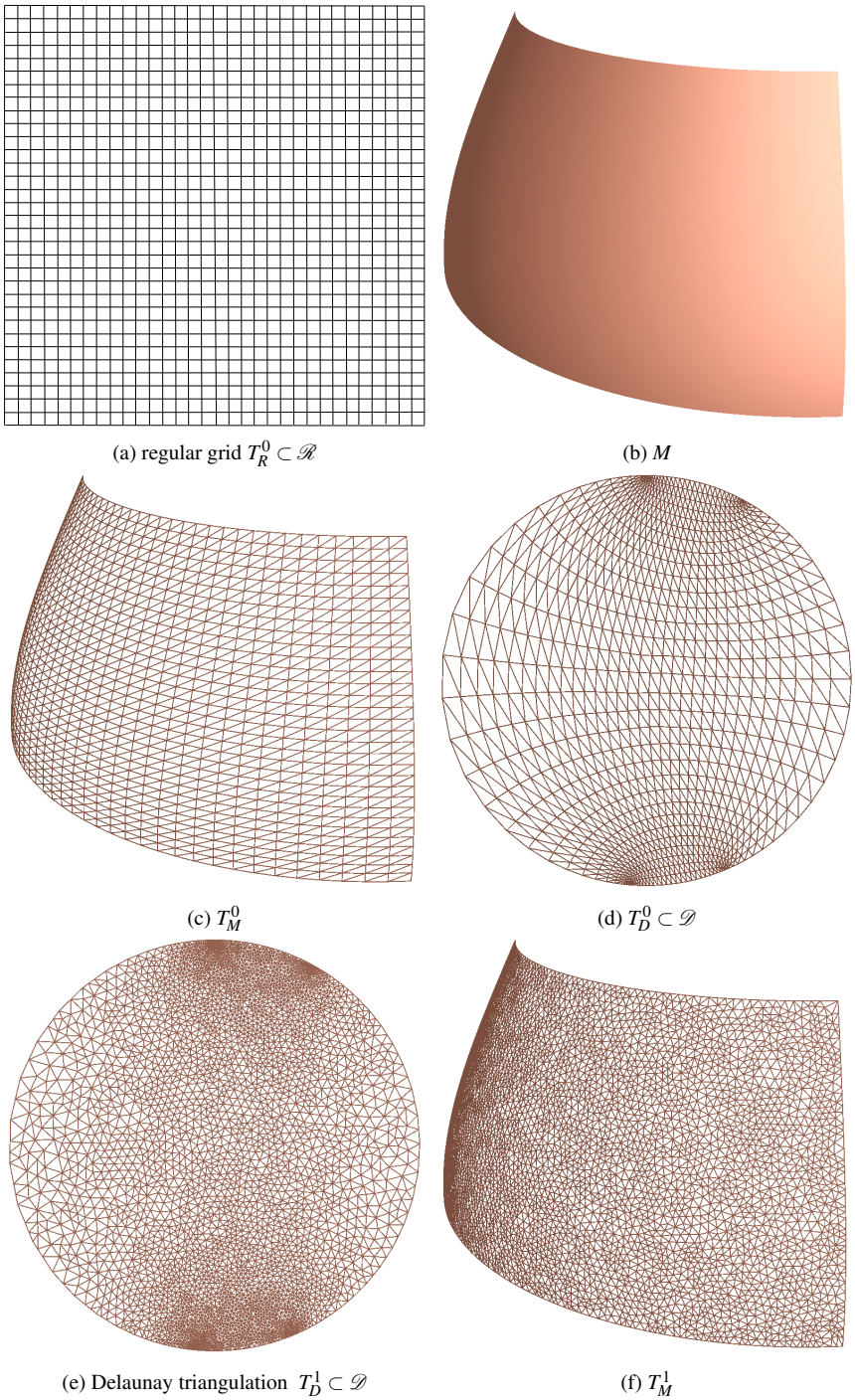
Remark The estimation for Gaussian curvature measure holds for surface intrinsically defined, namely, with a Riemannian metric but without embedding. We can prove it in the same way by applying Nash's embedding theorem [Nash 1954] [Nash 1956] and normal cycle theory which holds in any codimension.

5.5 Computational Algorithm

We verified our theoretical results by meshing spline surfaces and comparing the Gaussian and mean curvature measures.

Each spline patch M is represented as a parametric smooth surface defined on a planar rectangle $\gamma : \mathscr{R} \to \mathbb{E}^3$, where $\mathscr{R}$ is the planar rectangle parameter domain, the position vector γ is C^2 continuous, therefore the classical curvatures are well defined. Let $\varphi : \mathscr{D} \to M$ be the conformal mapping from the unit disk $\mathscr{D}$ to the spline surface M. As shown in the left-hand diagram in Diagram (5.20), the mapping f is from $\mathscr{D}$ to $\mathscr{R}$, which makes the diagram commute, therefore $f = \gamma^{-1} \circ \varphi$.

$$\begin{array}{ccc} \mathscr{R} & \xrightarrow{\gamma} & M \\ {\scriptstyle f}\uparrow & \nearrow_{\varphi} & \\ \mathscr{D} & & \end{array} \qquad \begin{array}{ccc} T_R^k & \xrightarrow{\gamma} & T_M^k \\ {\scriptstyle \bar{f}}\uparrow & \nearrow_{\bar{\varphi}} & \\ T_D^k & & \end{array} \tag{5.20}$$



(a) regular grid $T_R^0 \subset \mathscr{R}$ (b) M

(c) T_M^0 (d) $T_D^0 \subset \mathscr{D}$

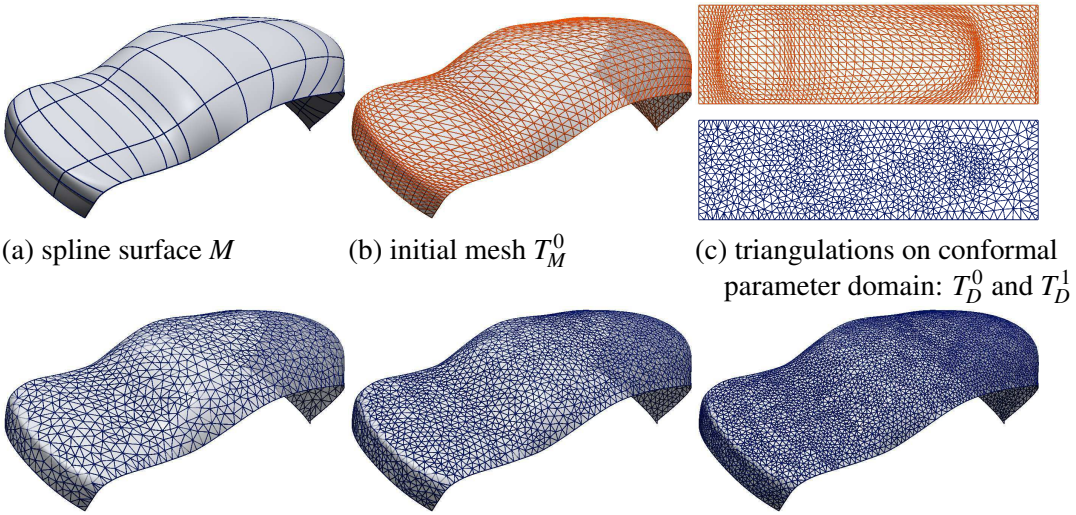
(e) Delaunay triangulation $T_D^1 \subset \mathscr{D}$ (f) T_M^1

Figure 5.8: *Pipeline for meshing a Bézier patch of Utah teapot.*

As shown in Fig. 5.8, in our experiments, each planar domain or surface S ($S \in \{\mathscr{D}, \mathscr{R}, M\}$), is approximated by two triangle meshes, $T_S^k, k = 0, 1$, where the T_S^0 is induced by the regular grid on the rectangle; T_S^1 is induced by the Delaunay triangulation on the unit disk. Both conformal parameterization φ and the parameter domain mapping

f are approximated by piecewise linear (PL) mappings, $\bar{\varphi}$ and $\bar{f}$, respectively, which are computed on the meshes.

5.5.1 Algorithm Pipeline



(a) spline surface M (b) initial mesh T_M^0 (c) triangulations on conformal parameter domain: T_D^0 and T_D^1

(d) Delaunay meshing result T_M^1 with $1K$ samples (e) Delaunay meshing result T_M^1 with $2K$ samples (f) Delaunay meshing result T_M^1 with $4K$ samples

Figure 5.9: *Remeshing of the Car spline surface model.*

5.5.1.1 Conformal Parametrization

In the first stage, the conformal parameterization is computed as follows:

$$\bar{f}^{-1}: \; T_R^0 \xrightarrow{\gamma} T_M^0 \xrightarrow{\bar{\varphi}^{-1}} T_D^0$$

T_R^0 is a triangulation induced by the regular grids structure on the rectangle $\mathscr{R}$. Each vertex on T_R^0 is mapped to the spline surface M by γ, each face is mapped to a Euclidean triangle, this gives the mesh T_M^0. If the grid tessellation is dense, the quality of the mesh T_M^0 is good enough for performing the Ricci flow and we get the PL mapping $\bar{\varphi}^{-1}$, which maps T_M^0 to a triangulation of the disk T_D^0. The composition of $\bar{\varphi}$ and γ^{-1} gives the PL mapping $\bar{f} = \gamma^{-1} \circ \bar{\varphi} : T_D^0 \to T_R^0$.

5.5.1.2 Resampling and Remeshing

The process in the second stage is described in the following diagram:

$$\bar{\varphi}: \; T_D^1 \xrightarrow{\bar{f}} T_R^1 \xrightarrow{\gamma} T_M^1$$

First, we apply Ruppert's Delaunay refinement method to generate the triangulation T_D^1 with good quality on the unit disk. The triangulation on the disk T_D^1 is mapped to a triangulation T_R^1 on the rectangle by the PL mapping $\bar{f} : T_D^0 \to T_R^0$. The connectivity of T_R^1 is the

same as that of T_D^1. The vertices of T_R^1 are the images of the vertices of T_D^1 under the PL mapping $\bar{f}$, which are calculated as follows. Suppose q is a Delaunay vertex of T_D^1 on the disk, covered by a triangle $[p_0, p_1, p_2] \in T_D^0$. Assume the barycentric coordinates of q are $(\alpha_0, \alpha_1, \alpha_2)$, $q = \sum_k \alpha_k p_k$, then

$$\bar{f}(q) = \sum_k \alpha_k \bar{f}(p_k).$$

The triangulation T_R^1 induces a triangle mesh T_M^1, whose connectivity is that of T_R^1, vertices of T_M^1 are the images of those of T_R^1 under the spline mapping γ. The discrete PL conformal mapping is given by

$$\bar{\varphi} = \gamma \circ \bar{f} : T_D^1 \to T_M^1.$$

The triangle mesh generated by the Delaunay refinement based on conformal parameterization is T_M^1.

Fig. 5.9 shows the meshing results using the proposed method for a car model. In this experiment, the conformal parameter domain $\mathscr{D}$ is also a rectangle. Frame (a) shows a B-spline surface patch M; Frame (b) shows the initial triangle mesh T_M^0; Frame (c) shows the triangualtions on the conformal parameter domain, $T_{\mathscr{D}}^0$ on the top and $T_{\mathscr{D}}^1$ at the bottom; Frames (d), (e) and (f) illustrate the triangle meshes generated by Delaunay refinement on conformal parameter domain with a different number of samples, $1K$, $2K$, and $4K$, respectively.

5.6 Experimental Results

Table 5.2: *The numbers of vertices and triangles of the sequence of meshes* $\{T_n\}$ *with different resolutions.*

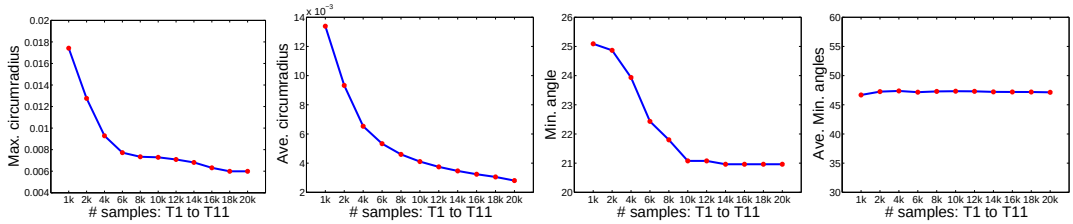
mesh	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}
# vertex	1021	2045	4093	6141	8189	10237	12285	14333	16381	18429	21876
# triangle	1904	3910	7950	11973	16040	20118	24192	28249	32316	36372	43202

The meshing algorithms are developed using generic C++ on Windows platform, all the experiments are conducted on a PC with Intel Core 2 CPU, 2.66GHz, 3,49G RAM.

5.6.1 Triangulation Quality

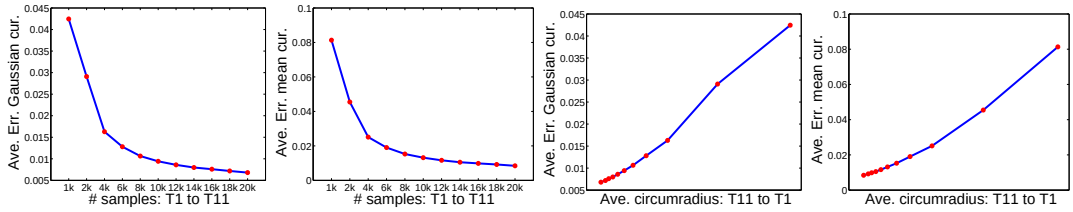
The patch on the Utah teapot (see Fig. 5.8) is meshed with different sampling densities, the meshes are denoted as $\{T_n\}_{n=1}^{11}$ as in Table 5.2. The statistics of the meshing quality are reported in Fig. 5.10. Frame (a) shows the maximal circumradius of all the triangles of each mesh. Frame (b) is the average circumradius of all the triangles of each mesh, because the sampling is uniform, we expect the circumradius ε_n vs. the number of vertices s_n to satisfy the relation

$$\varepsilon_n \sim \frac{1}{\sqrt{s_n}}.$$



(a) s_n vs. Max. ε_n (b) s_n vs. Ave. ε_n (c) s_n vs. Min. angle (d) s_n vs. Ave. Min. angles

Figure 5.10: *The maximal and average circumradii* $\{\varepsilon_n\}$ *(a-b), and the minimal and average of minimal angles of* $\{T_n\}$ *(c-d).*



(a) s_n vs. e_n^G (b) s_n vs. e_n^H (c) Ave. ε_n vs. e_n^G (d) Ave. ε_n vs. e_n^H

Figure 5.11: *Curvature errors* e_n^G *and* e_n^H *of* $\{T_n\}$ *converge to zeros as the number of sample points goes to infinity (a-b), and as the average of circumradii* $\{\varepsilon_n\}$ *goes to zero (c-d).*

The curve in Frame (b) perfectly meets our expectations. Frames (c) and (d) show the minimal angles on all meshes. According to the theory of Rupert's Delaunay refinement, the minimal angle should be no less than 20.7°. Frame (c) shows the minimal angles; in our experiments they are no less than 20.9°. Frame (d) illustrates the means of the minimal angles, which exceed 46.5°.

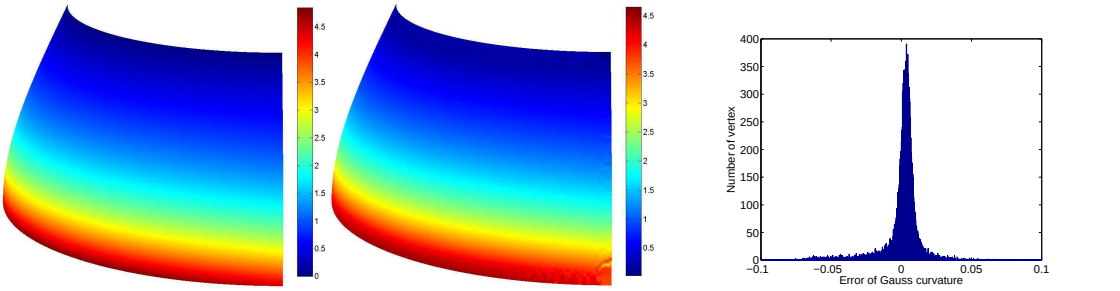
5.6.2 Curvature Measure Comparisons

For each triangle mesh T_k produced by our method, for each vertex $\mathbf{q} \in T_k$, we define a small ball in $\mathbb{R}^3$, $B(\mathbf{q},r)$ centered at $\mathbf{q}$ with radius r. We then calculate the curvature measures $\phi_{T_k}^G(B(\mathbf{q},r))$ and $\phi_{T_k}^H(B(\mathbf{q},r))$ using the formulae Eqn. 5.1 and Eqn. 5.2, respectively.

We also compute the curvature measures on the smooth surface M, $\phi_M^G(B(\mathbf{q},r))$ and $\phi_M^H(B(\mathbf{q},r))$ using the following method,

$$\phi_M^G(B(\mathbf{q},r)) := \int_{\gamma(u,v)\in B(\mathbf{q},r)} G(u,v)g(u,v)dudv,$$

where $\gamma(u,v)$ is the point on the spline surface. The $G(u,v)$ is the Gaussian curvature at $\gamma(u,v)$, $g(u,v)$ is the determinant of the metric tensor. Because the spline surface is C^2 continuous, all the differential geometric quantities can be directly computed using traditional formulae. Note that, because M and T_k are very close, we use $B(\mathbf{q},r) \cap T_k$ to replace $\pi(B(\mathbf{q},r)) \cap M$ in practice. In all our experiments, we set r to be $0.05area(M)^{\frac{1}{2}}$ and $0.08area(M)^{\frac{1}{2}}$ for Gaussian and mean curvature measures, respectively.



(a) Gaussian curvature of M (b) Gaussian curvature of T_{11} (c) histogram of Gaussian curvature error

(d) mean curvature of M (e) mean curvature of T_{11} (f) histogram of mean curvature error

Figure 5.12: *Illustration of the curvature values on the Utah teapot spline surface patch M, (a, d), and on its approximate mesh T_{11} (b, e). Their relative curvature error distribution histograms are shown in (c) and (f).*

We define the average error between curvature measures as

$$e_n^G = \frac{1}{|V_n|} \sum_{v \in V_n} |\phi_M^G(B(v,r)) - \phi_{T_n}^G(B(v,r))|,$$

and

$$e_n^H = \frac{1}{|V_n|} \sum_{v \in V_n} |\phi_M^H(B(v,r)) - \phi_{T_n}^H(B(v,r))|,$$

where V_n is the vertex set of T_n.

Figure 5.11 shows the errors between curvature measures with respect to sampling densities, or equivalently, the number of samples and the average circumradius. Frames (a) and (b) show the curvature measure error is approximately proportional to the inverse of the square root of the number of sample points; Frames (c) and (d) show the curvature measure error is approximately linear with respect to the circumradius. This again matches our main Theorem 5.3.2.

Figure 5.12 visualizes the curvature distributions on the smooth patch M (left column), and the triangle mesh T_{11} (right column). The histograms show the distributions of the relative curvature errors at the vertices of the mesh. From the two left-hand columns, we can see that the curvatures of M look very similar to their counterparts on T_{11}. Moreover, from the right-hand column, we can find that the overwhelming majority of vertices have relative curvature errors very close to zeros. In particular, for Gaussian curvature measure, more than 97% vertices are gathered in the relative error range of (-0.05, 0.05). For mean

curvature measure, more than 95% of vertices are in the relative error range of (-0.05,0.05). This demonstrates the accuracy of the proposed method.

5.7 Conclusion

This work analyzes the surface meshing algorithm based on conformal parameterization and the Delaunay refinement method. Based on the normal cycle theory and conformal geometry, we rigorously prove the curvature measure convergence and estimate Hausdorff distance and normal deviation. According to [Hildebrandt *et al.* 2005], these theoretical results also imply the convergence of Riemannian metric and Laplace-Beltrami operator.

The method can be generalized to prove the curvature convergence of other meshing algorithms, such as the centroidal voronoi tessellation method, and so on. The normal cycle theory is general to arbitrary dimension. We will generalize the theoretical results of this work to higher dimensional discretizations, such as volumetric shapes. We will explore these directions in the future.

CHAPTER 6

Conclusions and Future Work

Contents

6.1 Contributions 171
6.1.1 3D Face Recognition in the Presence of Various Expressions 171
6.1.2 3D Face Recognition in the Presence of Expression, Pose and Occlusion Variations 172
6.1.3 3D Data Quality Improvement: Surface Meshing with Curvature Convergence 172
6.2 Perspectives for Future Work 173
6.2.1 3D Face Recognition: Encoding Surface Normal or Coordinate? . . 173
6.2.2 3D Partial Face Recognition with Data Quality Guarantee 173
6.2.3 3D Video-based Face Recognition 174

This Ph.D. thesis mainly concentrates on the topic of 3D face recognition in the real. In chapter 1, we give a general introduction of the contexts, motivations, objectives, methodologies, and contributions of this thesis. In chapter 2, we extensively review three categories of 3D face recognition approaches: *i.e.* expression deformation modeling based, surface-distance based, and local region/feature based. Our main research works are introduced in the following three chapters. In Chapter 3, we present a 3D face recognition approach to handle various expressions. and In chapter 4, we display a registration free local feature-based approach to handle expression, pose and occlusion issues in an uniform framework. In chapter 5, we transfer our attention to improve the data-quality of 3D mesh. In the last chapter, we concludes this thesis and list some suggestions for the future work.

6.1 Contributions

The contributions in this thesis are as follows.

6.1.1 3D Face Recognition in the Presence of Various Expressions

To handle expression variations, we present an effective and general approach by using a novel highly discriminative facial shape descriptor along with a facial shape deformation learning strategy. The proposed facial shape descriptor is based on three facial normal images encoded by multi-scale local binary patterns and enables accurate and fast characterization of local shape changes. It is much more discriminative than both raw normal images and range images encoded in the same way. The facial shape deformation learning strategy is accomplished by running the Sparse Representation-based Classifier (SRC),

which computes the recognition scores of local facial regions at a given normal image and encoding scale. This approach achieves a rank-one score up to 96.3% on the FRGC v2.0 database. Furthermore, experimental results on the Bosphorus, BU-3DFE and 3D-TEC databases demonstrate a good generalizability with respect to various types of facial expressions. Finally, the proposed approach is also very fast. It takes only 3.55 seconds to identify a single probe in a gallery of 466 subjects such as those in the FRGC v2.0 database.

6.1.2 3D Face Recognition in the Presence of Expression, Pose and Occlusion Variations

To handle expression, pose and occlusion in an uniform framework, we present a mesh-based registration free 3D face recognition approach using novel pose invariant local shape descriptors and a multi-task sparse representation-based fine grained face matching. Facial salient vertices are automatically located using the minimum and maximum curvatures through a 3D Gaussian scale space. The local shape of each salient vertex is comprehensively described by histograms of multiple-order surface differential quantities. In addition to a Coarse Grained Matching scheme (CGM) which simply counts the number of matched vertices, we also propose a Fine Grained Matching scheme (FGM), where the similarity between a pair of face scans is measured as the average spare representation reconstruction error. Comprehensive experiments are carried out on *Bosphorus*, the largest public database which best accounts for the various challenges of 3D FR in uncontrolled conditions, *i.e.* expression variations, pose changes and external occlusions. The experimental results demonstrate the effectiveness of the proposed method which achieves the best state of the art rank-1 recognition rates over the whole dataset (96.56%) and its three subsets of expression variations (98.82%), pose changes (91.14%) and occlusions (99.21%). Additional experiments tested on the *FRGC v2.0* database show that the proposed approach also has a good generalizability.

6.1.3 3D Data Quality Improvement: Surface Meshing with Curvature Convergence

Surface differential quantities such as surface normal and curvatures are the basic building blocks of the proposed facial surface descriptors in this thesis. Thus, their accurate estimations play a key role for reliable facial surface descriptions. To this aim, we present a practical surface meshing algorithm which can largely improve the mesh quality of 3D data and thereby accurately estimate the surface differential quantities. Given a metric surface, the proposed algorithm triangulates its conformal uniformization domain by planar Delaunay refinement, and produces high quality meshes. Furthermore, we give explicit estimates for the Hausdorff distance, normal deviation, and differences of curvature measures between the surface and the mesh. Our strong estimates directly guarantee the convergence of Hausdorff distance, normal deviation, and curvature measure. Meanwhile, our result on Gaussian curvature measure is intrinsic to the Riemannian metric and independent of the embedding. Finally, the experimental results verified the theoretical results and demonstrated the efficiency of the approach.

6.2 Perspectives for Future Work

In this section, we present some possible extensions and future directions of this thesis.

6.2.1 3D Face Recognition: Encoding Surface Normal or Coordinate?

Facial surface normal information, as the first order surface differential quantity, determines the orientation of the surface at each point and contains informative local surface shape information. However, the majority of existing 3D face recognition approaches directly compare the point-distance between facial surfaces (*e.g.* ICP-bsed), or extract facial shape descriptors directly from the geometric coordinates of range image or depth image. Recently, normal information has been successfully utilized for facial surface description [Kakadiaris *et al.* 2007a] and facial surface registration [Queirolo *et al.* 2010] [Mohammadzade & Hatzinakos 2013]. As an important conclusion in [Mohammadzade & Hatzinakos 2013], the authors show that the surface normal vectors of the face at the sampled points contain more discriminatory information than the coordinates of the points. In chapter 3, we also use Local Binary Patterns (LBP) to encode the original normal information and indicates that the encoded normal information (LNP) is much more discriminative than the encoded depth information (LSP). It seems that the superiority of using surface normal for 3D face recognition becomes more and more clear. Thus, we believe that this conclusion is true in general.

It will be a very valuable research work if one can design a general and fair comparison scheme to comprehensively compare the performance of local facial shape descriptors extracted based on the surface normal information and the surface depth (coordinates) information. As a preliminary extension, in [Li *et al.* 2012], we also propose two new ways to encode the normal information, including multi-scale and multi-orientation Gabor Filters (GFs) and Local Gabor Binary Patterns (LGBPs). Together with MSMC-LNPs, we achieve a group of facial normal descriptors consisting of Normal Local Binary Patterns descriptor (N-LBPs, i.e. MSMC-LNPs), Normal Gabor Filters descriptor (N-GFs) and Normal Local Gabor Binary Patterns descriptor (N-LGBPs). The effectiveness of the proposed descriptors are evaluated on the 3D TEC dataset for recognizing identical twins, and we find that all these descriptors are much more discriminative than those counterpart ones based on depth images.

6.2.2 3D Partial Face Recognition with Data Quality Guarantee

3D partial face recognition is a common scenario for real-life biometric applications and pose a great challenge to current face recognition approaches. The challenge is mainly come from: 1. data missing caused by pose variations, and 2. low-quality data (*e.g.* low-resolution, irregular sampling and irregular triangles). The data missing issue has been studied in several works, see *e.g.* Kakadiaris *et al.* [Perakis *et al.* 2009] [Passalis *et al.* 2011] [Chu *et al.* 2013], Maes *et al.* [Maes *et al.* 2010], Li *et al.* [Li *et al.* 2011c], Berretti *et al.* [Berretti *et al.* 2011b] [Berretti *et al.* 2013], Drira *et al.* [Drira *et al.* 2013]. However, the low-quality data issue is still open. As mentioned in chapter 1, when we say that a facial triangle mesh is of "low-quality", either it has a sparse set of sampling points or it has a large set of triangles with irregular shapes, or both. All these cases can make the computation of

multiple order differential quantities highly unreliable. In chapter 5, we introduce a surface meshing algorithm to generate "high-quality" surface mesh, from which the convergence of surface normal and curvatures can be guaranteed. Based on the result of this algorithm, an immediate extension of the work in chapter 4 is that how large of the performance improvement for 3D partial face recognition can be achieved after we change the low-quality face data to high-quality one. We leave this work in near future.

6.2.3 3D Video-based Face Recognition

In the last several decades, face recognition in 2D video has received significant attention [Liu & Cheng 2003] [Lee *et al.* 2003], see [Wang *et al.* 2009] for a comprehensive survey. Not only since the informative dynamic information contained in video sequence has a possibility to largely improve the still image-based recognition performance, but also the wide range of real-life biometric applications (*e.g.* video surveillance). We believe that similar things will be happened in the field of 3D face recognition in the near future. along with the development of 3D shape scanning, tracking, and description techniques. For example, two published 3D video-based face databases have already been used in the study of 3D dynamic facial expression recognition (the BU-3DFE Database [Yin *et al.* 2008b]) and 3D gender recognition (the EURECOM Kinect Face Dataset [Huynh *et al.* 2013]). And the 3D surface registration and tracking techniques are also becomes a very active research topic in recent years [Zeng *et al.* 2010b] [Zeng *et al.* 2011]. Combing all these achievements and the temporal extensions of the local facial shape descriptors proposed in this thesis, it is possible for us to study the new topic of 3D video-based face recognition, which will be a new direction of real-life 3D face recognition in the near future.

APPENDIX A

Smooth Surfaces

In this chapter, we will give a short introduction of the basic concepts and theorems of smooth surface. Locally, the geometry of a smooth surface patch (i.e., a simple surface) can be better described and characterized based on its parametrization. For example, based on a parametric representation of a smooth surface patch, we can define its first fundamental form and second fundamental form. And then, the concepts of area, normal and curvatures can be well defined and computed. Meanwhile, the structure of a smooth surface can also be simply characterized by using the tool of moving frame (i.e., differential forms).

Furthermore, to describe the global geometry information of a complex topology surface, the concept of local coordinate chart and atlas play a fundamental effect. Based on them, we can define smooth manifold, conformal structure and Riemann surface. Meanwhile, we can analysis the mappings between two Riemann surfaces, such as conformal mapping, smooth map and tangent map. Then, the Riemannian metric on a smooth manifold can be defined from the product tangent space to a real number. Any Riemann surface endowed with a Riemannian metric composes a Riemannian surface. By generalizing the concept of Gaussian curvature and theorem of Gauss-Bonnet to Riemannian surface, we can define the surface Ricci flow on a Riemannian surface. The uniformization theorem tell us that the universal covering space of a surface can be isometrically embedded onto one of the three canonical spaces: the sphere, the Euclidean plane or the hyperbolic disk. Finally, we give the definition of the degree of a map between two manifolds. We can see that the orientation and the topology of a surface is very related to the degree of the map defined on it.

The main contents of this chapter are written by referring to the Lecture Notes and the Tutorials of Prof. Xianfeng David Gu. They are "Lecture Notes on Computational Conformal Geometry","Tutorial on Riemann Surface", "Tutorial on Euclidean Ricci flow" etc. [1] Meanwhile, the following books: [Shifrin 2010], [Carmo 1976], [Mei 2003] and [Bai 1992] have been our valuable supplementary references.

A.1 Local Differential Geometry

A.1.1 Regular parameterized Surface Patch

Suppose $D = \{(u,v)\}$ is a planar domain, a map $\mathbf{r} : D \to \mathbb{R}^3$,

$$\mathbf{r}(u,v) = (x(u,v), y(u,v), z(u,v)),$$

satisfying

[1] All these Lecture Notes and the Tutorials are free available on Prof. Xianfeng David Gu's homepage: http://www.cs.sunysb.edu/ gu/.

- $x(u,v), y(u,v), z(u,v)$ are differentiable of class C^2;
- $\mathbf{r}_u$ and $\mathbf{r}_v$ are linear independent, namely

$$\mathbf{r}_u = (\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}), \mathbf{r}_v = (\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v}), \mathbf{r}_u \times \mathbf{r}_v \neq 0,$$

is called a C^2 smooth regular surface patch in $\mathbb{R}^3$ (for C^2 map $\mathbf{r}$, we have $\mathbf{r}_{uv} = \mathbf{r}_{vu}$), (u,v) are the coordinates parameters of the surface $\mathbf{r}(u,v)$.

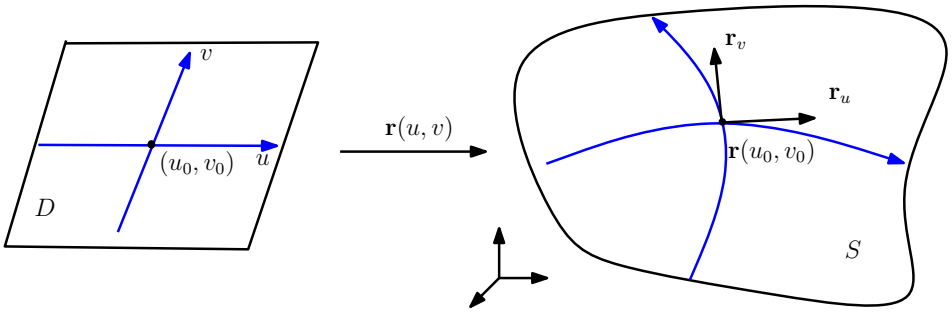


Figure A.1: *Regular parameterized surface patch S.*

A.1.2 First Fundamental Form and Area

Définition 10. *Given a surface S in $\mathbb{R}^3$ with parametric representation $\mathbf{r} = \mathbf{r}(u,v)$. The quadratic differential from*

$$\begin{aligned} I &= \langle d\mathbf{r}, d\mathbf{r} \rangle = \langle \mathbf{r}_u du + \mathbf{r}_v dv, \mathbf{r}_u du + \mathbf{r}_v dv \rangle \\ &= \langle \mathbf{r}_u, \mathbf{r}_u \rangle du^2 + 2\langle \mathbf{r}_u, \mathbf{r}_v \rangle du dv + \langle \mathbf{r}_v, \mathbf{r}_v \rangle dv^2 \end{aligned}$$

is called the first fundamental form of S.

In particular, if denote

$$E = \langle \mathbf{r}_u, \mathbf{r}_u \rangle, F = \langle \mathbf{r}_u, \mathbf{r}_v \rangle, G = \langle \mathbf{r}_v, \mathbf{r}_v \rangle,$$

then

$$I = d\mathbf{r}^2 = ds^2 = Edu^2 + 2Fdudv + Gdv^2.$$

Définition 11. *Given a surface S in $\mathbb{R}^3$ with parametric representation $\mathbf{r} = \mathbf{r}(u,v)$. Then, its area element*

$$d\sigma = \|\mathbf{r}_u du \times \mathbf{r}_v dv\| = \|\mathbf{r}_u \times \mathbf{r}_v\| dudv.$$

According to the Lagrange Equation: $(\mathbf{x} \times \mathbf{y})^2 = \mathbf{x}^2\mathbf{y}^2 - (\mathbf{x}\mathbf{y})^2$, we have

$$\|\mathbf{r}_u \times \mathbf{r}_v\|^2 = \mathbf{r}_u^2 \mathbf{r}_v^2 - (\mathbf{r}_u \mathbf{x}_v)^2 = EG - F^2 > 0,$$

Thus, the area element $d\sigma$ of S is also equals to

$$d\sigma = \sqrt{EG - F^2} dudv.$$

A.1.3 Normal, Second Fundamental Form and Curvatures

Définition 12. *Suppose a surface S has a parametric representation* $\mathbf{r} = \mathbf{r}(u,v)$, $\mathbf{r}_u$ *and* $\mathbf{r}_v$ *are coordinate tangent vector of S, then the unit normal vector of S is*

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\|\mathbf{r}_u \times \mathbf{r}_v\|}.$$

Définition 13. *The the second fundamental form of S is defined as*

$$\begin{aligned} II &= -\langle d\mathbf{r}, d\mathbf{n}\rangle = \langle \mathbf{r}_u du + \mathbf{r}_v dv, \mathbf{n}_u du + \mathbf{n}_v dv\rangle \\ &= \langle \mathbf{r}_u, \mathbf{n}_u\rangle du^2 + \langle \mathbf{r}_u, \mathbf{n}_v\rangle dudv + \langle \mathbf{r}_v, \mathbf{n}_u\rangle dudv + \langle \mathbf{r}_v, \mathbf{n}_v\rangle dv^2. \end{aligned}$$

Since

$$\begin{aligned} \mathbf{r}_u \cdot \mathbf{n} = 0 &\Rightarrow (\mathbf{r}_u \cdot \mathbf{n})_v = \mathbf{r}_{uv} \cdot \mathbf{n} + \mathbf{r}_u \cdot \mathbf{n}_v = 0, \\ \mathbf{r}_u \cdot \mathbf{n} = 0 &\Rightarrow (\mathbf{r}_u \cdot \mathbf{n})_u = \mathbf{r}_{uu} \cdot \mathbf{n} + \mathbf{r}_u \cdot \mathbf{n}_u = 0, \\ \mathbf{r}_v \cdot \mathbf{n} = 0 &\Rightarrow (\mathbf{r}_v \cdot \mathbf{n})_u = \mathbf{r}_{vu} \cdot \mathbf{n} + \mathbf{r}_v \cdot \mathbf{n}_u = 0, \\ \mathbf{r}_v \cdot \mathbf{n} = 0 &\Rightarrow (\mathbf{r}_v \cdot \mathbf{n})_v = \mathbf{r}_{vv} \cdot \mathbf{n} + \mathbf{r}_v \cdot \mathbf{n}_v = 0, \end{aligned}$$

We have

$$\begin{aligned} \mathbf{r}_u \cdot \mathbf{n}_v &= -\mathbf{r}_{uv} \cdot \mathbf{n} = -\mathbf{r}_{vu} \cdot \mathbf{n} = \mathbf{r}_v \cdot \mathbf{n}_u, \\ \mathbf{r}_u \cdot \mathbf{n}_u &= -\mathbf{r}_{uu} \cdot \mathbf{n}, \\ \mathbf{r}_v \cdot \mathbf{n}_v &= -\mathbf{r}_{vv} \cdot \mathbf{n}. \end{aligned}$$

Define functions,

$$\begin{aligned} L &= -\langle \mathbf{r}_u, \mathbf{n}_u\rangle = \langle \mathbf{r}_{uu}, \mathbf{n}\rangle \\ M &= -\langle \mathbf{r}_u, \mathbf{n}_v\rangle = -\langle \mathbf{r}_v, \mathbf{n}_u\rangle = \langle \mathbf{r}_{uv}, \mathbf{n}\rangle \\ N &= -\langle \mathbf{r}_v, \mathbf{n}_v\rangle = \langle \mathbf{r}_{vv}, \mathbf{n}\rangle \end{aligned}$$

Then the second fundamental form is represented as

$$II = -\langle d\mathbf{r}, d\mathbf{n}\rangle = \langle d^2\mathbf{r}, \mathbf{n}\rangle = Ldu^2 + 2Mdudv + Ndv^2.$$

The second fundamental form characterizes how to bend of a surface near a point, it approximately equals to two times of the distance between surface and its tangent plane at a given point. As shown in Fig. A.2,

The position vectors of P and P' are $\mathbf{r}(s)$ and $\mathbf{r}(s+\triangle s)$. By Taylor formula, we have

$$\overrightarrow{PP'} = \mathbf{r}(s+\Delta s) - \mathbf{r}(s) = \dot{\mathbf{r}}\Delta s + \frac{1}{2}(\ddot{\mathbf{r}} + \varepsilon)(\Delta s)^2,$$

where $\lim_{\Delta s \to 0} \varepsilon = \mathbf{0}$. the distance between P' and π:

$$\delta = \overrightarrow{PP'} \cdot \mathbf{n} = [\mathbf{r}(s+\Delta s) - \mathbf{r}(s)] \cdot \mathbf{n} = \frac{1}{2}(\ddot{\mathbf{r}} \cdot \mathbf{n} + \varepsilon \cdot \mathbf{n})(\Delta s)^2.$$

Thus, when $\ddot{\mathbf{r}} \cdot \mathbf{n} \neq 0$,

$$2\delta \approx \mathbf{n} \cdot \ddot{\mathbf{r}}(\Delta s)^2 = \langle d^2\mathbf{r}, \mathbf{n}\rangle = II.$$

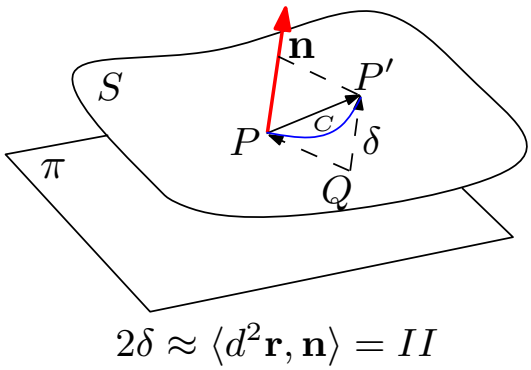


Figure A.2: *Illustration of the second fundamental form.*

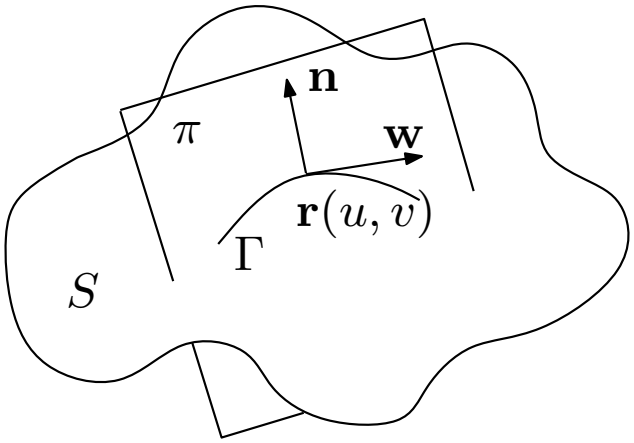


Figure A.3: *Normal curvature.*

Définition 14. *(Normal curvature) Suppose* $\mathbf{w} = a\mathbf{r}_u + b\mathbf{r}_v$ *is a tangent vector of S at point* $\mathbf{r}(u,v)$*, a plane* π *through normal* $\mathbf{n}$ *and* $\mathbf{w}$*, the planar curve* $\Gamma = S \cap \pi$ *has curvature* k_n*, at point* $\mathbf{r}(u,v)$*, which is called the normal curvature of S along the tangent vector* $\mathbf{w}$*. See Fig. A.3.*

Suppose $\mathbf{w} = a\mathbf{r}_u + b\mathbf{r}_v$ is a tangent vector of S at point $\mathbf{r}(u,v)$, the normal curvature along $\mathbf{w}$ is

$$k_n(\mathbf{w}) = \frac{II(\mathbf{w},\mathbf{w})}{I(\mathbf{w},\mathbf{w})}.$$

On convex surface patch, the normal curvature along any directions are positive. On the saddle surface patch, the normal curvature may be positive and negative, or zero (See Fig. A.4).

Définition 15. *(Gauss map) Suppose S is a surface with parametric representation* $\mathbf{r}(u,v)$*, the normal vector at point* (u,v) *is* $\mathbf{n}(u,v)$*, the mapping*

$$\mathbf{g} : S \to S^2, \mathbf{r}(u,v) \to \mathbf{n}(u,v),$$

is called the Gauss map of S, (see Fig. A.5).

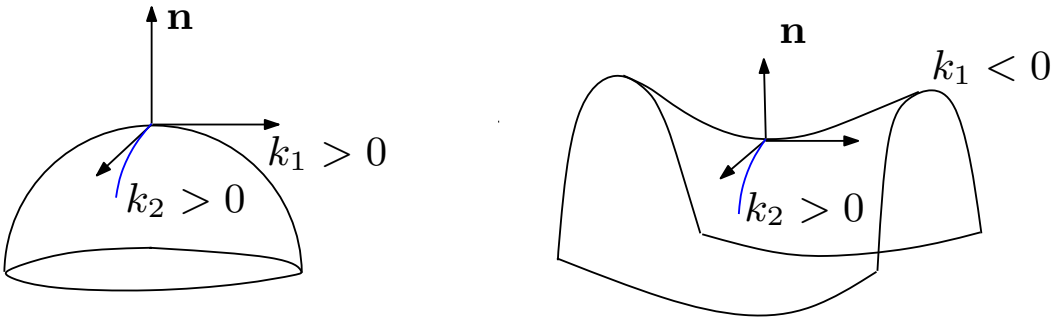


Figure A.4: *Normal curvature.*

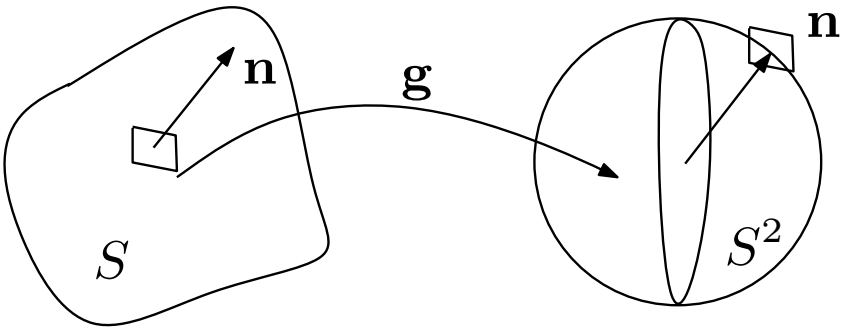


Figure A.5: *Gauss map.*

Définition 16. *(Weingarten transform) The differential map* $\mathbf{W}$ *of Gauss map* $\mathbf{g}$ *is called the Weingarten transform.* $\mathbf{W}$ *is a linear map from the tangent space of S to the tangent space of* S^2*,*

$$\mathbf{W} : T_pS \to T_pS^2$$
$$\mathbf{v} = a\mathbf{r}_u + b\mathbf{r}_v \to \mathbf{W}(\mathbf{v}) = -(a\mathbf{n}_u + b\mathbf{n}_v).$$

The Weingarten transform has the following properties:

- Weingarten transform is independent of the choice of the parameters;
- Suppose $\mathbf{v}$ is a unit tangent vector of S, the normal curvature
$$k_n(\mathbf{v}) = \langle \mathbf{W}(\mathbf{v}), \mathbf{v} \rangle;$$
- Weingarten transform is self-conjugate transform from tangent plane to itself.
$$\langle \mathbf{W}(\mathbf{v}), \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{W}(\mathbf{v}) \rangle.$$

Définition 17. *(Principle curvature) The eigenvalues of Weingarten transformation is called* principle curvatures*. The Corresponding eigenvectors are called* principle directions*, namely*

$$\mathbf{W}(\mathbf{e_1}) = k_1\mathbf{e_1}, \mathbf{W}(\mathbf{e_2}) = k_1\mathbf{e_2},$$

where $\mathbf{e_1}$ *and* $\mathbf{e_2}$ *are unit vectors. Since the Weingarten transformation is self conjugate, it is symmetric. Therefore, the principle directions are orthogonal.*

Proposition A.1.1. *(Euler's Formula) Suppose an arbitrary unit tangent vector* $\mathbf{v} = \cos\theta e_1 + \sin\theta e_2$*, then the normal curvature along* $\mathbf{v}$ *is*

$$k_n(\mathbf{v}) = \langle \mathbf{W}(\mathbf{v}), \mathbf{v} \rangle = \cos^2\theta k_1 + \sin^2\theta k_2,$$

therefore, normal curvature reaches the maximum and minimum at the principle curvatures.

Proposition A.1.2. *The matrix representing of the Weingarten map* $\mathbf{W}$ *with respect to the basis* $\{\mathbf{r}_u, \mathbf{r}_v\}$ *is*

$$W = \frac{II}{I} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \cdot \begin{bmatrix} L & M \\ M & N \end{bmatrix},$$

In particular, if $\{\mathbf{r}_u, \mathbf{r}_v\}$ *is an orthogonal basis for* T_pS*, then* W *equals to* II*.*

Définition 18. *The product of the principle curvatures is called the* Gaussian curvature*:* $K = detW = k_1k_2$*. The average of the principle curvatures is called the* mean curvature*:* $H = \frac{1}{2}trW = \frac{1}{2}(k_1 + k_2)$*. We say* S *is a* minimal surface *if* $H = 0$ *and* flat *if* $K = 0$*.*

Corollary A.1.3.

$$K = k_1k_2 = \frac{LN - M^2}{EG - F^2}, \quad H = \frac{1}{2}(k_1 + k_2) = \frac{LG - 2MF + NE}{2(EG - F^2)},$$

In particular, if there is no umbilical points and the parameter curves are lines of curvature, then $F = M = 0$*, and* $k_1 = \frac{L}{E}$, $k_2 = \frac{N}{G}$*. Then,* $K = \frac{LN}{EG}$, $H = \frac{LG+NE}{2EG}$*.*

A.1.4 Surface Theory with Differential Forms

Suppose S is a surface with parametric representation $\mathbf{r}(u, v)$. we define a *moving frame* $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ on surface S. Here $\mathbf{e_i}$ are vector fields defined on S with the properties that:

- $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ gives an orthogonal basis at each point;
- $\mathbf{e_1}, \mathbf{e_2}$ is a basis for the tangent space of S and $\mathbf{e_3} = \mathbf{n}$.

Then, we define 1-forms ω_1 and ω_2 on S by

$$d\mathbf{r} = \omega_1\mathbf{e_1} + \omega_2\mathbf{e_2};$$

Then, we have the first fundamental form

$$I = \langle d\mathbf{r}, d\mathbf{r} \rangle = \omega_1^2 + \omega_2^2.$$

To see how the frame itself twists, we define 1-forms $\omega_{ij}, i, j = 1, 2, 3$, by

$$d\mathbf{e_i} = \sum_{j=1}^{3} \omega_{ij}\mathbf{e_j}.$$

Since $\mathbf{e_i} \cdot \mathbf{e_j} = 0$ for any $i, j = 1, 2, 3$, we have

$$0 = d(\mathbf{e_i} \cdot \mathbf{e_j}) = d(\mathbf{e_i}) \cdot \mathbf{e_j} + \mathbf{e_i} \cdot d(\mathbf{e_j}) = (\sum_{k=1}^{3} \omega_{ik}\mathbf{e_k}) \cdot \mathbf{e_j} + (\sum_{k=1}^{3} \omega_{jk}\mathbf{e_k}) \cdot \mathbf{e_i} = \omega_{ij} + \omega_{ji},$$

So $\omega_{ij} = \omega_{ji}$ for all i, $j = 1,2,3$, and $\omega_{ii} = 0$ for all i. i.e.,

$$\begin{pmatrix} de_1 \\ de_2 \\ de_3 \end{pmatrix} = \begin{pmatrix} 0 & \omega_{12} & \omega_{13} \\ -\omega_{12} & 0 & \omega_{23} \\ -\omega_{13} & -\omega_{23} & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix},$$

In particular, we have

$$d\mathbf{n} = d\mathbf{e_3} = \omega_{31}\mathbf{e_1} + \omega_{32}\mathbf{e_2} = -(\omega_{13}\mathbf{e_1} + \omega_{23}\mathbf{e_2}).$$

Thus, we have the second fundamental form

$$II = -\langle d\mathbf{r}, d\mathbf{n}\rangle = \omega_1\omega_{13} + \omega_2\omega_{23}.$$

When we write

$$\omega_{13} = h_{11}\omega_1 + h_{12}\omega_2$$
$$\omega_{23} = h_{21}\omega_1 + h_{22}\omega_2$$

for appropriate coefficient $h_{\alpha\beta}$, we can see that the Weingarten map $\mathbf{W}$ with respect to the basis $\{\mathbf{e_1}, \mathbf{e_2}\}$ for T_pS is nothing but $[h_{\alpha\beta}]$. According to the properties of the exterior derivative, we have

$$\begin{aligned} \mathbf{0} = d(d\mathbf{r}) &= d(\omega_1\mathbf{e_1} + \omega_2\mathbf{e_2}) = d\omega_1\mathbf{e_1} + d\omega_2\mathbf{e_2} - \omega_1 \wedge (\sum_{j=1}^{3} \omega_{1j}\mathbf{e_j}) - \omega_2 \wedge (\sum_{j=1}^{3} \omega_{2j}\mathbf{e_j}) \\ &= (d\omega_1 - \omega_2 \wedge \omega_{21})\mathbf{e_1} + (d\omega_2 - \omega_1 \wedge \omega_{12})\mathbf{e_2} - (\omega_1 \wedge \omega_{13} + \omega_2 \wedge \omega_{23})\mathbf{e_3}, \end{aligned}$$

From the third term, we have

$$0 = \omega_1 \wedge \omega_{13} + \omega_2 \wedge \omega_{23} = \omega_1 \wedge (h_{11}\omega_1 + h_{12}\omega_2) + \omega_2 \wedge (h_{21}\omega_1 + h_{22}\omega_2) = (h_{12} - h_{21})\omega_1 \wedge \omega_2,$$

Thus, we have $h_{12} = h_{21}$. That is to say, the Weingarten map is symmetric.

In particular, if $\mathbf{e_1}$ and $\mathbf{e_2}$ are principle directions if and only if $h_{12} = h_{21} = 0$, and we have $\omega_{13} = k_1\omega_1$ and $\omega_{23} = k_2\omega_2$, where k_1 and k_2 are principle curvatures. Based on these results, we can define the curvature measures as follows,

$$\omega_A = \omega_1 \wedge \omega_2 \tag{A.1}$$
$$\omega_K = \omega_{13} \wedge \omega_{23} = K\omega_1 \wedge \omega_2 \tag{A.2}$$
$$\omega_H = \omega_1 \wedge \omega_{23} - \omega_2 \wedge \omega_{13} = 2H\omega_1 \wedge \omega_2 \tag{A.3}$$

where ω_A (Eqn.A.1), ω_K (Eqn.A.2) and ω_H (Eqn.A.3) are the *area element*, the *Gauss curvature measure* and the *mean curvature measure*.

A.2 Riemann Surface: Global Differential Geometry

A.2.1 Riemann Surface and conformal mapping

Définition 19. *(Manifold) Suppose S is a topological surface. $\{U_\alpha\}, \alpha \in I$ is a open covering of S, $S \subset \cup_\alpha U_\alpha$. For each U_α, $\phi_\alpha : U_\alpha \to \mathbb{R}^n$ is a homeomorphism. The pair (U_α, ϕ_α) is a local coordinate chart. Suppose $U_\alpha \bigcap U_\beta \neq \emptyset$, the transition function*

$$\phi_{\alpha\beta} = \phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\beta \cap U_\alpha).$$

is smooth, then S is called a smooth manifold, $\{(U_\alpha, \phi_\alpha)\}$ is called an atlas.

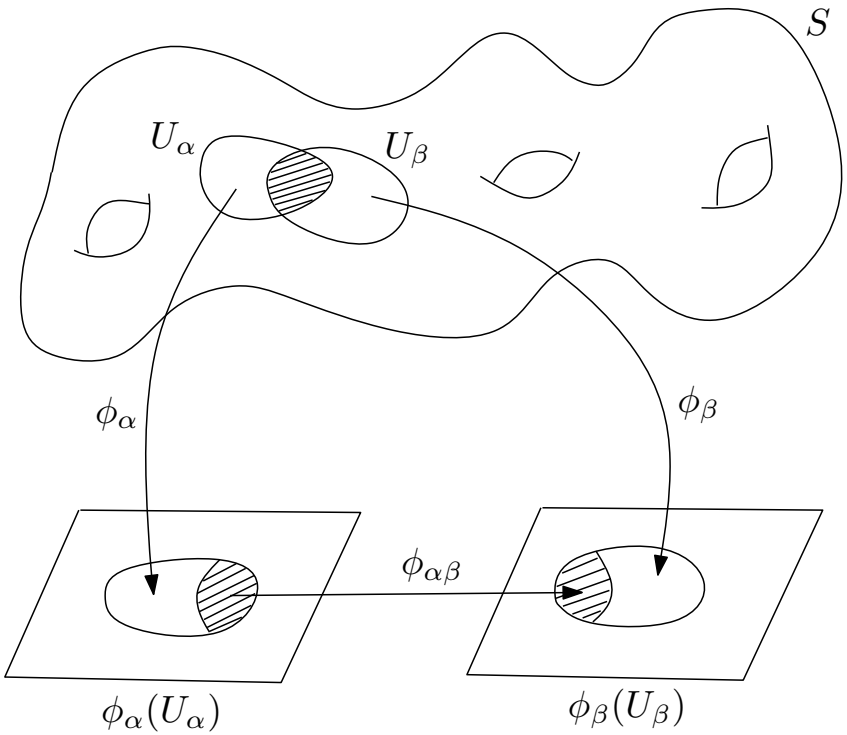


Figure A.6: *Smooth manifold.*

Définition 20. *(Holomorphic function) Let U be an open subset of $\mathbb{C}$, $f : U \to \mathbb{C}$ be a complex-valued function. We write, for all real numbers x an y, $f(x+iy) = u(x,y)+iv(x,y)$, where u denotes the real part of f and v its imaginary part. If f satisfies the following Cauchy-Riemann equations:*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x},$$

then f is a holomorphic *function. If both f and f^{-1} are holomorphic,then f is a bi-holomorphic function.*

Définition 21. *(Conformal atlas) Suppose S is a topological surface, (2-manifold), A is an atlas, such that all the chart transition functions $\phi_{\alpha\beta} : \mathbb{C} \to \mathbb{C}$ are bi-holomorphic, then A is called a conformal atlas. If B is another conformal atlas, and $A \cup B$ is still a conformal atlas, we say A and B are compatible.*

Définition 22. *(Conformal structure) Suppose S is a topological surface, consider all the conformal atlas on S, classified by the compatible relation*

$$\{\textit{all conformal atlas}\}/\sim,$$

each equivalence calss is called a conformal structure.

Définition 23. *(Riemann surface) A topological surface with conformal structure is called a* Riemann surface*.*

Définition 24. *(Conformal mapping) Let S_1 and S_2 be two Riemann surfaces with conformal structures $\{(U_\alpha, \phi_\alpha)\}$ and $\{(V_\beta, \eta_\beta)\}$. A map $f : S_1 \to S_2$ is* conformal*, if any local presentation*

$$\eta_\beta \circ f \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha) \to \eta_\beta(V_\beta)$$

is holomorphic.

Définition 25. *(Smooth map) Suppose $f : S_1 \to S_2$ is a map between two smooth manifolds. For each point p, choose a chart of S_1, $(U_\alpha, \phi_\alpha), p \in U_\alpha$. The image $f(U_\alpha) \subset V_\beta, (V_\beta, \tau_\beta)$ is a chart of S_2. The local representation of f*

$$\tau_\beta \circ f \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha) \to \tau_\beta(V_\beta)$$

is smooth, then f is called a smooth map.

A curve on a manifold is a map $\gamma : [0,1] \to M, \gamma(t) \in M$. Choose a local chart (U_α, ϕ_α) with local parameter (x,y), then the curve can be represented as $(x(t), y(t))$. The velocity vector of the curve is represented as

$$\frac{d\gamma(t)}{dt} = \frac{\partial}{\partial x}\frac{dx}{dt} + \frac{\partial}{\partial y}\frac{dy}{dt}.$$

Let $f : M \to N$ be a smooth map, then $f \circ \gamma : [0,1] \to N$ is a curve on N. Choose a local chart on N, (V_β, τ_β) with local parameters (u,v). Then the local representation of the map

$$\tau_\beta \circ f \circ \phi_\alpha^{-1} : (x,y) \to (u(x,y), v(x,y)),$$

the local representation of $f \circ \gamma$ is $(u(x(t), y(t)), v(x(t), y(t)))$.

The velocity of $f \circ \gamma$ is

$$\frac{f \circ \gamma}{dt} = \frac{\partial}{\partial u}\frac{du}{dt} + \frac{\partial}{\partial v}\frac{dv}{dt}.$$

where

$$\begin{bmatrix} du \\ dv \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix}.$$

Définition 26. *(Tangent map) The linear map $df : T_pM \to T_{f(p)}N$, which maps a tangent vector in T_pM to a tangent vector $T_f(p)N$,*

$$\frac{d\gamma(t)}{dt} \to \frac{f \circ \gamma}{dt}$$

is call the tangent map of f, or the push-forward map.

we say $\frac{f\circ\gamma}{dt}$ is the push-forward of $\frac{d\gamma(t)}{dt}$, and denote it as

$$f_*(\frac{d\gamma(t)}{dt}) = \frac{f \circ \gamma}{dt}.$$

The local representation of the tangent map is the Jacobi matrix.

Définition 27. *(Riemannian metric) A Riemannian metric on a smooth manifold M is an assignment of an inner product $\mathbf{g}_p : T_pM \times T_pM \to \mathbb{R}, \forall p \in M$, such that*

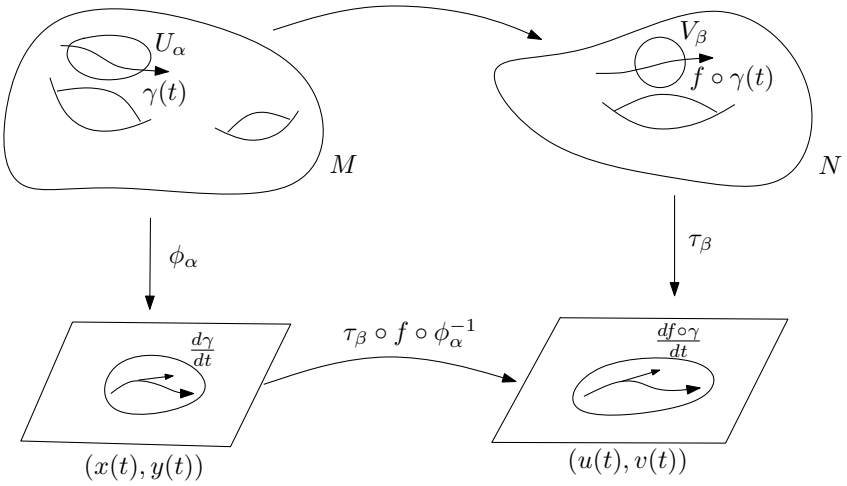


Figure A.7: *Tangent map.*

- $\mathbf{g}_p(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) = \sum_{i,j=1}^{2} a_i b_j g_p(X_i, Y_j)$;
- $\mathbf{g}(X,Y) = \mathbf{g}(Y,X)$;
- $\mathbf{g}$ *is non-degenerate;*
- $\forall p \in M$, *there exists local coordinates* $\{x_i, x_j\}$, *such that* $\mathbf{g}_{ij} = \mathbf{g}_p(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j})$ *are* C^∞ *functions.*

Définition 28. *(Pull back Riemannian metric) Suppose* $f : (M, \mathbf{g}) \to (N, \mathbf{h})$ *is smooth mapping between two Riemannian manifolds,* $\forall p \in M, f_* : T_pM \to T_f(p)N$ *is the tangent map. The pull back metric* $f^*\mathbf{h}$ *induced by the mapping* f *is given by*

$$f^*\mathbf{h}(X_1, X_2) := \mathbf{h}(f_*X_1, f_*X_2), \forall X_1, X_2 \in T_pM.$$

Local representation of the pull back metric is given by

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{pmatrix} \begin{pmatrix} \mathbf{h}_{11} & \mathbf{h}_{12} \\ \mathbf{h}_{21} & \mathbf{h}_{22} \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

Définition 29. *(Conformal equivalent metrics) Suppose* $\mathbf{g}_1$, $\mathbf{g}_2$ *are two Riemannian metrics on a manifold M, if*

$$\mathbf{g}_1 = e^{2u}\mathbf{g}_2, u : M \to \mathbb{R}$$

then $\mathbf{g}_1$ *and* $\mathbf{g}_2$ *are conformal equivalent.*

Définition 30. *(Conformal structure) Consider all Riemannian metrics on a topological surface S, which are classified by the conformal equivalence relation,*

$$\{\textit{Riemannian metrics on } S\}/\sim$$

each equivalence class is called a conformal structure.

Définition 31. *(Riemannian surface) Any Riemann surface may be endowed with a Riemannian metric* $\mathbf{g}$*, compatible with its holomorphic structure, that is, such that on each local coordinates chart with local parameters* (U_α, ϕ_α)*,* $\mathbf{g}$ *has the following expression :*

$$\mathbf{g} = e^{2\lambda(u_\alpha, v_\alpha)}(du_\alpha^2 + dv_\alpha^2),$$

where λ *is a smooth function. The couple* $(S, \mathbf{g})$ *is called a Riemannian surface. Such local coordinates are called* isothermal coordinates*.*

Conversely, on any metric surface $(S, \mathbf{g})$, the existence of isothermal coordinates. Then one can use isothermal coordinates to build an atlas, which gives a conformal structure of the surface. Therefore all Riemannian surfaces are Riemann surfaces.

Définition 32. *(Conformal mapping) If* S_1 *(resp.* S_2*) is endowed with a metric* $\mathbf{g}_1$ *(resp.* $\mathbf{g}_2$*) compatible with its holomorphic structure, then a map* $f : S_1 \to S_2$ *is conformal, if and only if the pull back metric induced by* f *satisfies*

$$f^*\mathbf{g}_2 = e^{2\lambda}\mathbf{g}_1.$$

All the conformal mappings from the unit disk to itself can be represented as a *Möbius transformation*,

$$z \to e^{i\theta}\frac{z - z_0}{1 - \bar{z}_0 z}, z_0, z \in \mathbb{D},$$

where $\mathbb{D}$ is the unit disk on the complex plane $|z| < 1$.

The Gaussian curvature of a Riemannian surface $(S, \mathbf{g})$ can be defined locally with help of isothermal coordinates as follows.

Définition 33. *(Gaussian Curvature) Let* $(S, \mathbf{g})$ *be a surface with a Riemannian metric* $\mathbf{g}$*, under the isothermal coordinates, the Gaussian curvature is defined as*

$$K(u, v) = -\frac{1}{e^{2\lambda(u,v)}}\Delta\lambda,$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

The operator $\frac{1}{e^{2\lambda(u,v)}}\Delta$ is called the Laplace-Beltrami operator of the surface. Gauss-Bonnet theorem states that the total curvature $\int_S K(p)dA$ of a closed surface S is a topological invariant.

Theorem A.2.1. *(Gauss-Bonnet) The total Gaussian curvature of* $(S, \mathbf{g})$ *equals* $2\pi\chi(S)$ *:*

$$\int_S K(p)dA = 2\pi\chi(S),$$

where $\chi(S)$ *is the Euler-characteristic of S.*

A.2.2 Ricci Flow

Suppose S is a surface endowed with two conformal Riemannian metrics $\mathbf{g}$ and $\bar{\mathbf{g}}$ conformal, $\bar{\mathbf{g}} = e^{2\lambda}\mathbf{g}$. Then the corresponding Gaussian curvature functions satisfy the following Yamabe equation

$$\bar{K} = e^{-2\lambda}(-\Delta_{\mathbf{g}}\lambda + K),$$

where $\Delta_{\mathbf{g}}$ is the Laplace-Beltrami operator induced by $\mathbf{g}$. Hamilton's Ricci flow is a powerful tool to solve the Yamabe equation.

Définition 34 (Hamilton's Surface Ricci Flow)**.** *Let S be a closed surface endowed with a Riemannian metric $\mathbf{g} = (g_{ij})$, the Ricci flow on it is the flow $g_{ij}(t)$ satisying the following equation*

$$\frac{dg_{ij}(t)}{dt} = -2K(t)g_{ij}(t), \tag{A.4}$$

where $K(t)$ is the Gaussian curvature of $g(t)$.

We can simplify this equation, setting $g(t) = e^{2u(t)}g(0)$. Then, normalizing A.4 so that the area of S is preserved during the deformation, the Ricci flow equation can be written

$$\frac{du(t)}{dt} = -2(K + \frac{2\pi\chi(S)}{A(t)}), \tag{A.5}$$

where $A(t)$ is the area under $\mathbf{g}(t)$.

Theorem A.2.2 (Hamilton and Chow)**.** *When t tends to $+\infty$, the Ricci flow A.5 converges, and induces a metric $\mathbf{g}(\infty)$ with constant Gaussian curvature.*

This metric $\mathbf{g}(\infty)$ is called the *uniformization metric*. Remark that at each t, $g(t)$ is conformal to $g(0)$. We apply surface Ricci flow for computing conformal mappings.

Définition 35 (Universal covering space)**.** *Let $p : \tilde{S} \to S$ be a continuous surjective map, such that each point $q \in S$ has a neighborhood U whose preimage $p^{-1}(U) = \bigcup_i \tilde{U}_i$ is a disjoint union of open sets $\tilde{U}_i$, and such that the restriction of p on each $\tilde{U}_i$ is a local homeomorphism. Then $(\tilde{S}, p)$ is a called a* covering space *of S, and p is called a projection map. If the covering space $\tilde{S}$ is simply connected, then $\tilde{S}$ is called the* universal covering space *of S.*

Theorem A.2.3 (Uniformization)**.** *Let $(S, \mathbf{g})$ be a compact Riemannian surface, then there exists a metric $\tilde{\mathbf{g}} = e^{2\lambda}\mathbf{g}$ on S conformal to $\mathbf{g}$, which has constant Gaussian curvature. Depending on the genus (g) of the surface, the constant curvature is $+1$ ($g = 0$), 0 ($g = 1$), or -1 ($g > 1$).*

Therefore the universal covering space of the surface can be isometrically embedded onto one of the three canonical spaces: the sphere $\mathbb{S}^2$, the Euclidean plane $\mathbb{E}^2$ or the hyperbolic disk $\mathbb{H}^2$. If the surface is with boundaries, then it can be conformally mapped onto circle domains in the canonical space, whose complements are geodesic circles.

A.2.3 Degree of Maps

Définition 36 (Degree of maps)**.** *Let (M,ω), (N,θ) be compact oriented manifolds of the same dimension, without boundaries. Assume N is connected. Let $f : M \to N$ be a C^1 map and $x \in M$ a regular point of f. Put $y = f(x)$. We say x has positive type if the isomorphism (tangent map) $T_x f : M_x \to N_y$ preserves orientation, that is, it sends ω_x to $theta_y$. In this case, we write $deg_x f = 1$. If $T_x f$ reverses orientation then x has negative type, and we write $deg_x f = -1$. We call $deg_x f$ the degree of f at x.*

Suppose $y \in N$ is any regular value for f. Define the degree of f over y to be

$$deg(f,y) = \sum_{x \in f^{-1}(y)} deg_x f;$$

if $f^{-1}(y)$ is empty, deg(f,y) = 0. To indicate orientation, we also write

$$deg(f,y) = deg(f,y;\omega,\theta).$$

Reversing ω, or θ changes the sign of $deg(f,y) = deg(f,y;\omega,\theta)$.

APPENDIX B

Discrete Surfaces

In this chapter, we will give a short introduction of the basic concepts and theorems of discrete surface corresponding to the ones of smooth surface introduced in Chapter 2. We begin with the concepts of simplex, simplicial complex, and triangular mesh. Then, we introduce definitions of the discrete normal and curvatures. Finally, we ended with the concept of discrete Riemannian metric and the theorem of discrete Euclidean Ricci flow.

The main contents of this chapter are written by referring to the Lecture Notes and the Tutorials of Prof. Xianfeng David Gu. They are "Lecture slides of Discrete surface", "Tutorial on Discrete Ricci flow" and "Tutorial on Euclidean Ricci flow" etc. [1]

B.1 Local Differential Geometry

B.1.1 Simplicial Complex and Triangular Mesh

Définition 37 (n-Simplex)**.** *Suppose $n+1$ points $\{v_0, v_1, \ldots, v_n\}$ in the general position in $\mathbb{R}^n$, the n-simplex $[v_0, v_1, \ldots, v_n]$ is the minimal convex set including all of them*

$$\sigma^n = [v_0, v_1, \ldots, v_n] := \{\sum_{i=0}^{n} \lambda_i v_i | \sum_{i=0}^{n} \lambda_i = 1, \lambda_i \geq 0\}$$

we call $v_0, v_1, \ldots, v_n$ the vertices of the n-simplex σ^n.

Fig. B.1 shows the examples of the 1, 2, 3-simplices in $\mathbb{R}^3$.

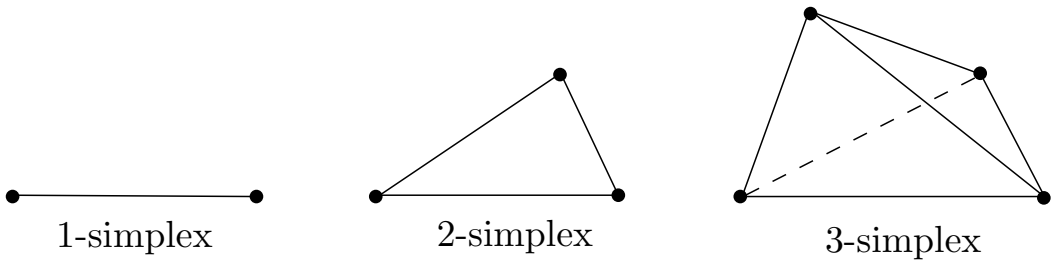


Figure B.1: *1, 2, 3-simplices in* $\mathbb{R}^3$.

Définition 38 (Face)**.** *Suppose σ^n is a simplex, $\tau^m \subset \sigma^n, (m < n)$, is also a simplex, then we say τ^m is a face of σ^n.*

[1] All these Lecture Notes and the Tutorials are free available on Prof. Xianfeng David Gu's homepage: http://www.cs.sunysb.edu/ gu/.

Définition 39 (Boundary Operator)**.** *The boundary of a n-simplex is*

$$\partial\sigma^n = \sum_{i=0}^{n}(-1)^j[v_0,\cdots,v_{j-1},v_{j+1},\cdots,v_n].$$

Définition 40 (Simplicial Complex)**.** *A simplicial complex Σ is a union of simplicies, such that*
1. If a simplex σ belongs to Σ, then all its faces also belongs to Σ.
2. If two simplicies $\sigma_1,\sigma_2 \subset \Sigma, \sigma_1 \cap \sigma_2 \neq \emptyset$, then their intersection is a common face.

A simplicial complex may not necessary to be a manifold. Fig. B.2 gives an example, where the simplicial complex is the union of a 1-simplex, a 2-simplex and a 3-simplex.

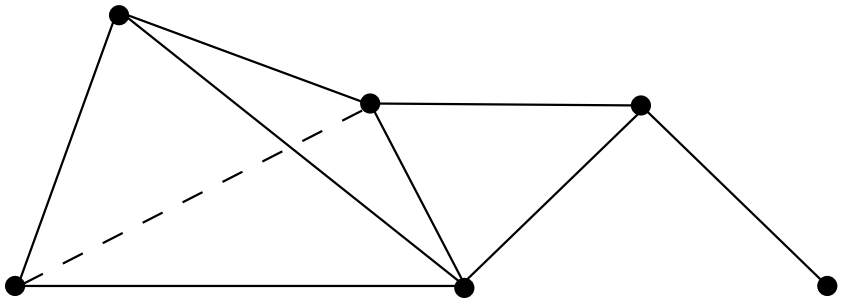


Figure B.2: *A non-manifold simplicial complex.*

Définition 41 (Triangular Mesh)**.** *Given a two dimensional simplicial complex Σ, if Σ is also a manifold, namely, for each point $p \in \Sigma$, there exists a neighborhood of p, $U(p)$, homeomorphic the whole plane or the upper half plane, then Σ is called a triangular mesh.*

If $U(p)$ is homeomorphic to the whole plane, then p is called a interior point; if $U(p)$ is homeomorphic to the upper half plane, then p is called a boundary point.

The combinatorics of the simplicial complex fully determines the topology of the surface. The geometric quantities, such as normal, curvature, can be regarded as functions defined on the edges and vertices of the complex. In the following discussion, we use $\Sigma = (V,E,F)$ to denote the mesh with vertex set V, edge set E and face set F.

B.1.2 Discrete Normal and Curvatures

Définition 42 (Discrete Normal)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh, $f = [v_0,v_1,v_2]$ is face of Σ. The discrete normal function $N : F \to \mathbb{R}^3$ defined on f is as follows,*

$$N(f) = \mathbf{e}_i \times \mathbf{e}_i,$$

where

$$\mathbf{e}_i = \frac{v_i - v_0}{|v_i - v_0|}.$$

Définition 43 (Discrete Gaussian Curvature)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh, the discrete Gaussian curvature function K is defined on vertices, $K : V \to \mathbb{R}^1$.*

$$K(v) = \begin{cases} 2\pi - \sum_i \alpha_i, & v \notin \partial M \\ \pi - \sum_i \alpha_i, & v \in \partial M. \end{cases} \tag{B.1}$$

where α_i's are corner angles adjacent to the vertex v, ∂M represents the boundary of the mesh.

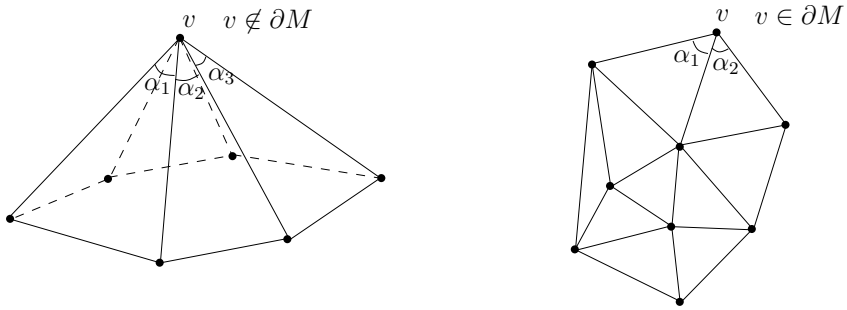


Figure B.3: *Gaussian curvature at point v.*

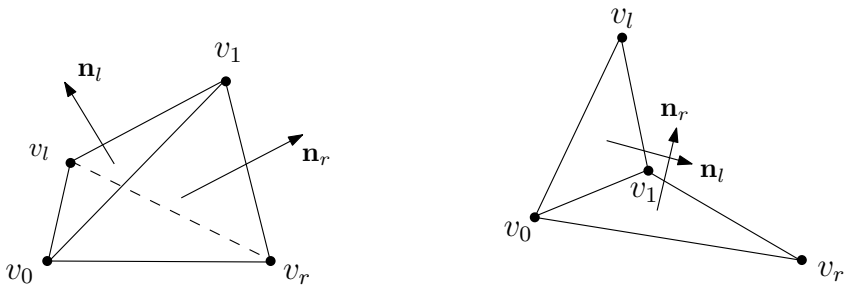
Définition 44 (Dihedral Angle)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh, suppose $e = [v_0,v_1] \in E$ be a edge on the mesh, with adjacent left face $[v_0,v_1,v_l]$ and the right face $[v_0,v_1,v_r]$. The normal of the left and the right faces are $\mathbf{n_l}$ and $\mathbf{n_r}$. Then the dihedral angle of the edge is given by*

$$\beta(e) = \cos^{-1} \mathbf{n_l} \times \mathbf{n_r} \cdot \mathbf{e},$$

where **e** *is the unit directional vector along the edge*

$$\mathbf{e} = \frac{v_1 - v_0}{|v_1 - v_0|}.$$

Note that, if the surface is convex, then $\beta(e)$ is positive, if the surface is concave, then $\beta(e)$ is negative (See Fig. B.4).



Positive dihedral angle Negetive dihedral angle

Figure B.4: *Dihedral angle on edge $e = [v_0,v_1]$.*

Définition 45 (Discrete Mean Curvature)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh, the discrete mean curvature function H is defined on edges, $H : E \to \mathbb{R}^1$.*

$$H(e) = l(e)\beta(e), e \notin \partial M.$$

where $l(e)$ represents the length of e and $\beta(e)$ represents the dihedral angle of e.

Theorem B.1.1 (Discrete Gauss-Bonnet Theorem)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh with Euclidean background metric, then the total curvature is a topological invariant,*

$$\sum_{v_i \notin V} K_i + \sum_{v_i \in V} K_i = 2\pi\chi(M).$$

B.2 Global Differential Geometry

B.2.1 Discrete Metric

Définition 46 (Discrete Riemannian Metric)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh, the discrete Riemannian metric on Σ is a function defined on the edges, $l : E \to \mathbb{R}^+$, satisfies triangular inequality: on each face $[v_i, v_j, v_k]$,*

$$l_{ij} + l_{jk} > l_{ki}, l_{jk} + l_{ki} > l_{ij}, l_{ki} + l_{ij} > l_{jk}.$$

The discrete metric determines the corner angles on each face by the cosine law,

$$\alpha_i = \cos^{-1} \frac{l_{ij}^2 + l_{ik}^2 - l_{jk}^2}{2 l_{ij} l_{ik}}.$$

Définition 47 (Circle Packing Metric)**.** *Let $\Sigma = (V,E,F)$ be a triangular mesh, we associate each vertex v_i with a circle with radius γ_i. On edge e_{ij}, the two circles intersect at the angle of ϕ_{ij}. The edge length are*

$$l(e_{ij}) = \gamma_i^2 + \gamma_j^2 + 2\gamma_i\gamma_j cos\phi_{ij}$$

The Circle Packing Metric is defined as (Σ, Γ, Φ), where $\Gamma = \{\gamma_i | \forall v_i\}$, $\Phi = \{\phi_{ij} | \forall e_{ij}\}$, (See Fig. B.5).

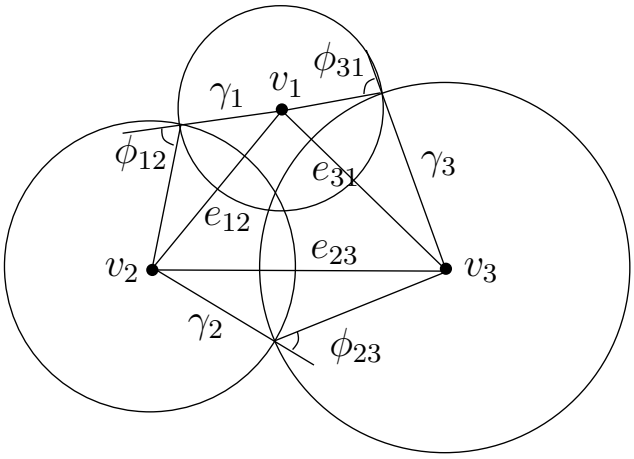


Figure B.5: *Circle packing metric.*

B.2.2 Discrete Euclidean Ricci Flow

Définition 48 (Conformal Factor)**.** *Define on each vertex* $\mathbf{u} : V \to \mathbb{R}$,

$$u_i = log\gamma_i.$$

Then, we have the following symmetric relation:

$$\frac{\partial K_i}{\partial u_j} = \frac{\partial K_j}{\partial u_i}.$$

Définition 49 (Discrete Ricci Flow)**.** *Given a circle packing metric* (Σ, Γ, Φ)*, the discrete Ricci flow is*

$$\frac{d\mathbf{u}}{dt} = \overline{\mathbf{K}} - \mathbf{K}$$

where $\mathbf{u} = (u_1, u_2, \cdots, u_n)$, $\mathbf{K} = (K_1, K_2, \cdots, K_n)$, $|V| = n$, *and* $\overline{\mathbf{K}} = (\overline{K_1}, \overline{K_2}, \cdots, \overline{K_n})$ *is the user defined target curvature.*

Définition 50 (Discrete Euclidean Ricci Energy)**.** *The discrete Euclidean Ricci energy is defined as*

$$E(\mathbf{u}) = \int_{\mathbf{u}_0}^{\mathbf{u}} \sum_i (\overline{K_i} - K_i) du_i,$$

where $\mathbf{u}_0 = (0, 0, \cdots, 0)$.

Theorem B.2.1 (Chow & Luo: Convexity of Euclidean Ricci Energy)**.** *The Euclidean Ricci energy* $E(\mathbf{u})$ *is strictly convex on the hyperplane* $\sum_i u_i = 0$.

APPENDIX C

Surface Discretization and Approximation

In this chapter, we transfer our attention to surface discretization and approximation. For surface discretization, we introduce the classical concept of Delaunay triangulation, meanwhile, one of the most famous Delaunay refinement algorithm, namely, Ruppert's Delaunay refinement algorithm is introduced. This algorithm plays a very important effect in the following chapter when we carry out surface meshing with curvature convergence.

On the surface geometry approximation theory, we first give two negative examples: the Lantern of Schwarz for area approximation and the angle defect for point-wise Gauss curvature approximation. From these two examples, we can find that with the condition of very closed Hausdorff distance between smooth surface and triangular mesh is not enough to guarantee a good approximation, more conditions related to the quality of the triangular mesh are also necessary. Based on these two examples, we introduce the concept of unified curvature measures (i.e., generalized curvatures), which is based on the normal cycle theory. Finally, the main approximation and convergence theorems of the curvature measures are introduced.

The contents of the surface approximation section is written by fully referring to the book "Generalized curvatures" written by Prof. Jean-Marie Morvan.

C.1 Surface Discretization

C.1.1 Delaunay Triangulation

Définition 51 (Delaunay Triangulation)**.** *Suppose Ω is a domain in $\mathbb{E}^2$. A sample point set $P = \{p_1, p_2, \cdots p_n\} \subset \Omega$ is in a general position (no more than three points in the same line and no more than four points in the same circle). a delaunay triangulation of P is a triangulation $DT(P)$ such that no point in P is inside the circumcircle of any triangle in $DT(P)$. Delaunay triangulations maximize the minimal angle of all the angles of the triangles in the triangulation; they tends to avoid skinny triangles. The Delaunay triangulations is unique if P is in general position.*

C.1.2 Ruppert's Delaunay Refinement Algorithm

In mesh generation, Ruppert's Delaunay Refinement Algorithm, proposed by Jim Ruppert in the early 1990s [Ruppert 1995], is an algorithm for creating 2D high quality Delaunay triangulations. The algorithm inputs a planar straight-line graph with non-acute angle and begins with a Delaunay triangulation of the input vertices. Then, it augments the vertices by the following two operations.

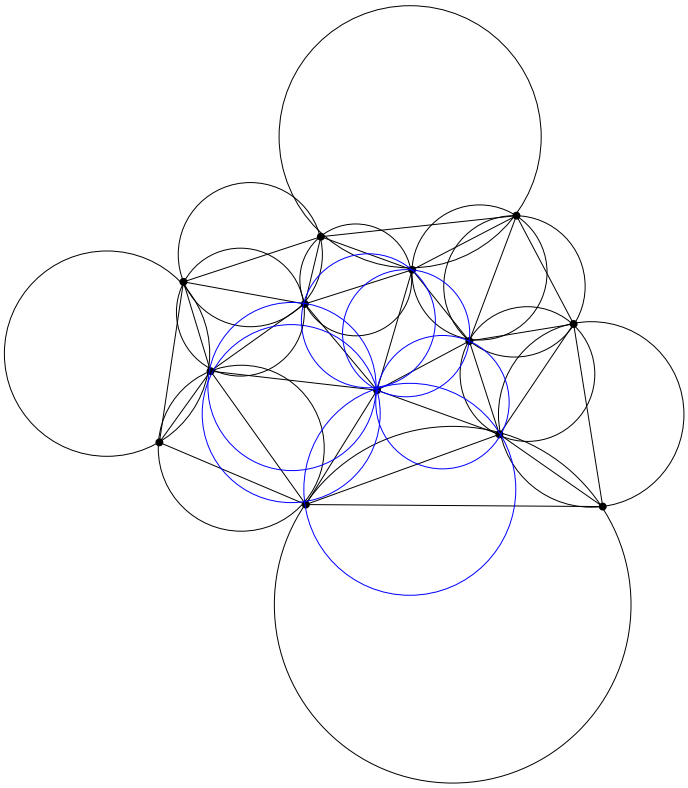


Figure C.1: *A Delaunay triangulation in* $\mathbb{E}^2$ *with circumcircles shown.*

1. Any encroached segment (a segment with non-empty diametral circles) is split into subsegments by inserting a new vertex at its midpoint, and then updating the Delaunay triangulation (See Fig. C.2).

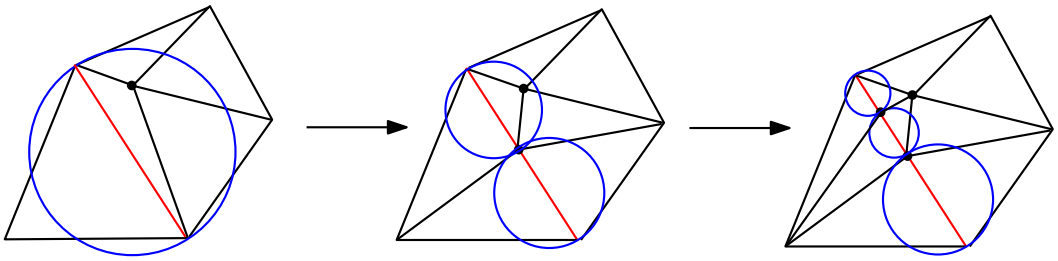


Figure C.2: *Any encroached segment is split into subsegments by inserting a new vertex at its midpoint.*

2. The circumcenter of a poor-quality triangle (a triangle with small angle) is insert into the triangulation, unless this circumcenter lies in the diametral circle of some segment. In this case, the encroached segment is split instead, and then updating the Delaunay triangulation (See Fig. C.3).

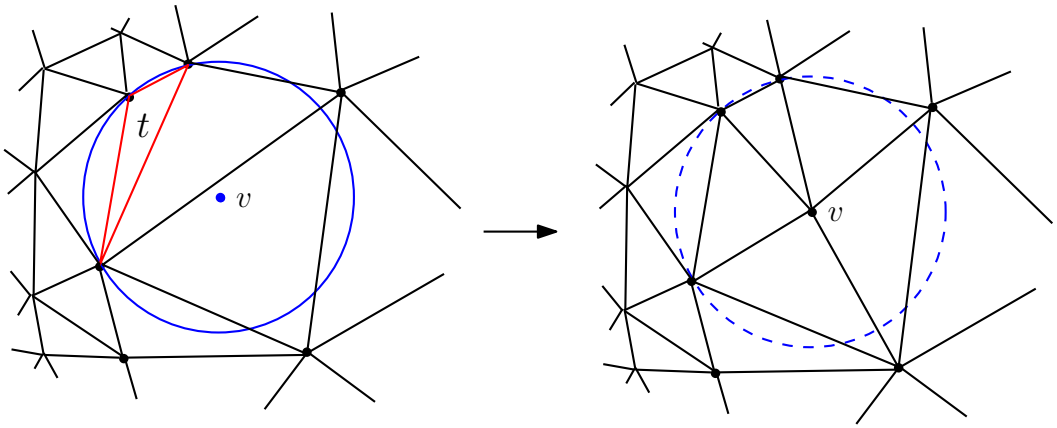


Figure C.3: *Insert vertices at circumcenters of triangles with small angles (e.g.,* $< 20.7°$*).*

The operations are repeated until no poor-quality triangles exist and all segments are not encroached. Finally, it returns the Delaunay triangulation of the augmented vertices.

The following theorems guarantee that Ruppert's algorithm generates high-quality triangulation.

Theorem C.1.1. *The triangulation produced by* Ruppert's *algorithm conforms to the input Planar straight line graph X and contains no angles less than* $20.7°$.

Définition 52 (Local feature size)**.** *Give a Planar straight line graph X, the* local feature size *of a point p, denoted* $lfs(p)$*, is the radius of the smallest disk centered at p that intersects two non-incident vertices or segments of X.*

Theorem C.1.2. *For* $\alpha < arcsin(\frac{1}{2\sqrt{2}}) \approx 20.7°$, Ruppert's *algorithm terminates. Moreover, there exists* $C_\alpha = \frac{1+\sqrt{2}}{1-2\sqrt{2}sin(\alpha)} > 0$ *such that for each vertex q inserted into the mesh,* $lfs(q) \leq C_\alpha r_q$*, where* $r_q = d(q,n_q)$ *and* n_q *is a nearest neighbor to q in the mesh.*

Theorem C.1.3. *Given a point p of the output mesh, its nearest neighbor point q is is a distance at least* $\frac{lfs(p)}{C_s+1}$*, where* $C_s = \frac{1+\sqrt{2}}{1-2\sqrt{2}sin(\alpha)}$*, for any* $\alpha < arcsin\frac{1}{2\sqrt{2}} \approx 20.7°$.

Theorem C.1.4. *For* $\alpha \leq 20.7°$*, Ruppert's algorithm terminates. Moreover, the number of points in the output mesh is most*

$$\sharp(points) \leq \frac{Area(B)}{C\varepsilon^2},$$

where B is the region enclosed by the bounding square, $\varepsilon = min\{lfs(x)|x \in B\}$*, and C is a constant to be specified.*

C.2 Surface Approximations

C.2.1 Two Negative Examples

Example 1. Area Approximation

One could think that the area of a smooth surface can be computed as the limit of the area of a sequence of triangulations "tending to it". Here we give a negative example, namely, the Lantern of Schwarz, to show that: one can construct a sequence of triangular meshes T_n inscribed in a smooth surface S in $\mathbb{E}^3$, moreover, the Hausdorff limit of T_n tends to S, however,

$$\lim_{n\to\infty} \mathbf{A}(T_n) \neq \lim \mathbf{A}(S).$$

Let C be a cylinder of finite height l and radius r in $\mathbb{E}^3$. Let $P(n,m)$ be a triangulation inscribed in C defined as follows: consider $m+1$ circles on the cylinder C obtained by intersecting C with 2-planes orthogonal to the axis of C. Inscribe in each circle a regular n-gon of the slice $k-1$ by a rotation of angle $\frac{\pi}{n}$. Then, join each vertex v of the slice $k-1$ to the two vertex of the slice k which are nearest to v. One obtains a triangulation whose vertices v_{ij} are defined as follows: for all $i \in \{0,\ldots,n-1\}$ and all $j \in \{0,\ldots,m\}$,

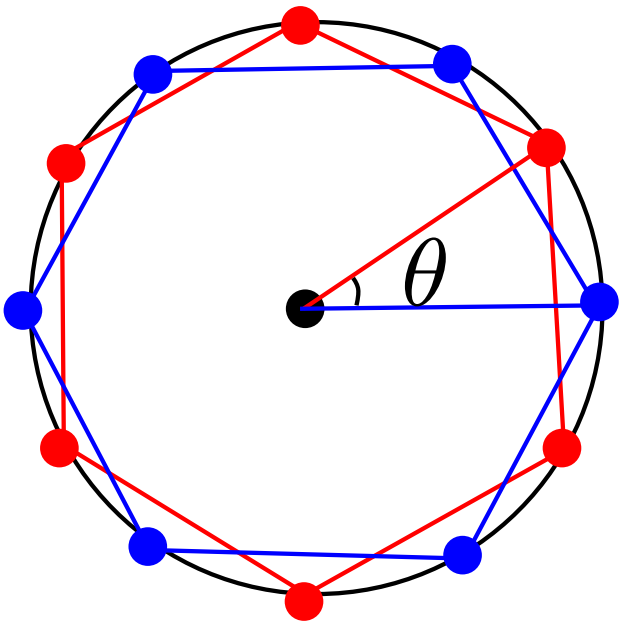


Figure C.4: *Hexagons inscribed in a circle,* $n = 6, \theta = \frac{\pi}{6}$.

$$v_{ij} = \begin{cases} (r\cos(i\alpha), r\sin(i\alpha), jh) & \text{if } j \text{ is even,} \\ (r\cos(i\alpha+\frac{\alpha}{2}), r\sin(i\alpha+\frac{\alpha}{2}), jh) & \text{if } j \text{ is odd,} \end{cases} \tag{C.1}$$

and whose faces are

$$v_{ij}v_{(i+1)j}v_{i(j+1)},$$
$$v_{ij}v_{(i-1)(j+1)}v_{i(j+1)},$$

where $\alpha = \frac{2\pi}{n}$ and $h = \frac{l}{m}$ (See Fig. C.4 and Fig. C.5 [1]).

Note that the area of the cylinder is $\mathbf{A}(C) = 2\pi rl$. On the other hand, the area $\mathbf{A}(P(n,m))$ of $P(n,m)$ is nothing but the sum of the areas of its triangles. When n tends to

[1] This figure is downloaded from the web of google image.

infinity, a simple computation shows that

$$\lim_{n\to\infty} \mathbf{A}(P(n,m)) = 2\pi rl\sqrt{1+\frac{m^2\pi^4 r^2}{4n^4 l^2}}.$$

In particular,

$$\lim_{n\to\infty} \mathbf{A}(P(n,n)) = 2\pi rl\sqrt{1+\frac{\pi^4 r^2}{4l^2}} \neq \mathbf{A}(C)$$

and

$$\lim_{n\to\infty} \mathbf{A}(P(n,n^3)) = +\infty,$$

although C is the Hausdorff limit of both $P(n,n^2)$ and $P(n,n^3)$.

Figure C.5: *Illustration of a Schwarz Lantern.*

This example shows that it is impossible to find sequence of triangulations inscribed in a (smooth) surface S whose Hausdorff limit is C, but whose area tends to infinity or to a limit different from the area of C.

Example 2. Gauss Curvature Approximation

A classical result on Gauss curvature approximation is that: the Gauss curvature at a point p of a smooth surface S can be recovered by the angle differ of flattening a *geodesic triangle* incident to p to *Euclidean geodesic triangle* [**?**]. As shown in Fig. C.6, let τ_i be a geodesic triangle on S with vertices p, p_i and p_{i+1}, denote by l_p, l_i, l_{i+1} the length of geodesic arcs opposite to the vertices. Finally, let β_i be the angle of τ_i at p. Now we flat τ_i to obtain corresponding *Euclidean geodesic triangles* t_i whose vertices are v, v_i, v_{i+1} in $\mathbb{E}^2$ and whose edge lengths satisfy

$$l_p = |v_i v_{i+1}|, l_i = |v v_{i+1}|, l_{i+1} = |v v_i|.$$

Let γ_i be the angle of t_i at v. Then the Gauss curvature K_p at p satisfy

$$\beta_i = \alpha_i + \frac{1}{6}K_p l_i l_{i+1} sin\alpha_i + o(l_{max}^2),$$

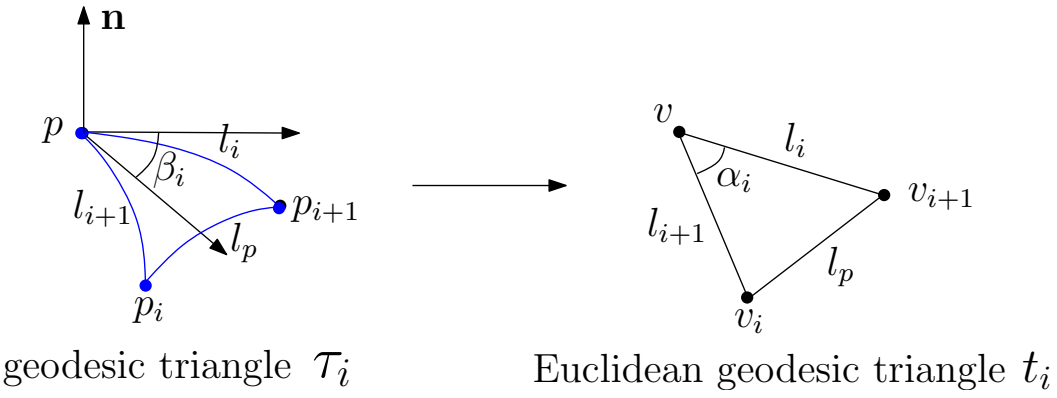


Figure C.6: *Flattening a geodesic triangle derives an estimation of the Gauss curvature.*

where $l_{max} = \max(l_i, l_{i+1})$.

When we consider all the geodesic triangles around p, its easy to see $\sum_i \beta_i = 2\pi$. Based on this point, one gets:

Theorem C.2.1.

$$2\pi - \sum_i \alpha_i = \frac{\sum_i \mathbf{A}(t_i)}{3} K_p + o(l_{max}^2).$$

Unfortunately, both the angle α_i and the area of t_i are rely on the information of the geodesic triangles τ_i, making this estimation unpractical. A nature insight is that when the edges lengths pp_i go to zero, one can approximate the t_i by directly replacing the geodesic arcs from p to p_is on S by the Euclidean segments pp_is. Indeed, this method is commonly used in practice, and is called *angle defect* point-wise Gauss curvature estimation method. See definition 43 in Chapter 3. However, the *angle defect* method is not reliable in most cases.

A deep study of the reliable of this method can be found in [Borrelli *et al.* 2003]. Here we give a short summarize of their main results. The basic idea to obtain the results is that performing the Taylor expansion of the angle defect in terms of the geometry of the surface around one of its points.

Consider a point p on a smooth surface S together with n Euclidean triangles $\{p_i p p_{i+1}\}$ (with $p_{n+1} = p_1$), forming a piecewise approximation of S around p, refer to Fig. we give the following notations.

- The 1-*ring* around p is the set of points $p_1, p_2, \ldots, p_n$ which have a common edge with p.
- The integer n is called the *valence* of p.
- Considering the plane Π_i, containing the normal $\mathbf{n}_p$, p and p_i. This plane is defined by its angle φ_i with respect to some coordinate system in the tangent plane of S at p.
- The Euclidean distance from p to p_i is denoted by η_i.
- The angel at p of the triangle pp_ip_{i+1} is denoted by γ_i.

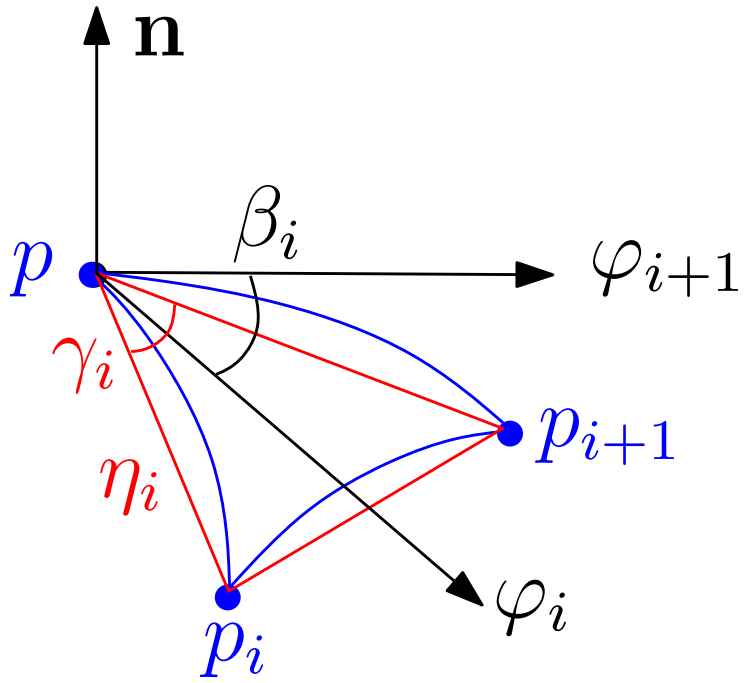


Figure C.7: *Notations.*

- The angle between two consecutive normal sections Π_i and Π_{i+1} is denoted by β_i.
- To simply the notations, we put

$$c_i = \cos\varphi_i, s_i = \sin\varphi_i.$$

Let

$$A_i = \frac{1}{4sin\gamma_i}[\eta_i\eta_{i+1}(c_i^2 s_{i+1}^2 + s_i^2 c_{i+1}^2) - \frac{\cos\gamma_i}{2}(\eta_i^2(2c_i^2 s_i^2) + \eta_{i+1}^2(2c_{i+1}^2 s_{i+1}^2))],$$
$$B_i = \frac{1}{4sin\gamma_i}[\eta_i\eta_{i+1}(c_i^2 c_{i+1}^2 - \frac{\cos\gamma_i}{2}(\eta_i^2 c_i^4 + \eta_{i+1}^2 c_{i+1}^4),$$
$$C_i = \frac{1}{4sin\gamma_i}[\eta_i\eta_{i+1}(s_i^2 s_{i+1}^2 - \frac{\cos\gamma_i}{2}(\eta_i^2 s_i^4 + \eta_{i+1}^2 s_{i+1}^4).$$

Let

$$A = \sum_i A_i, B = \sum_i B_i, C = \sum_i C_i.$$

One has the following theorem.

Theorem C.2.2. *At each point p of S,*

$$2\pi - \sum_{i=1}^{n} \gamma_i = (AK_p + B\lambda_1^2 + C\lambda_2^2) + o(\eta^2),$$

where λ_1 and λ_2 denotes the principle curvature of S at p.

In full generality, this theorem shows that there is no hope to get any meaningful approximation of the Gauss curvature of the surface, up to order 2, with the angle defect method.

The only possibility is to find special triangulations around p which cancel the terms B and C. This happens for the triangulations having some kind of regularity. let's introducing the following definition.

Définition 53. *Let p be a point of a smooth surface S and p_i, $i = 1,2,\ldots,n$ be its 1-*ring *neighbors. The point p is called a regular vertex if: (i) $\forall i, \beta_i = \frac{2\pi}{n}$; (ii) $\forall i, \eta_i = \eta$.*

In particular, we have the following theorem,

Theorem C.2.3. *Consider a regular vertex p with valence 6. Then*

$$\lim_{\eta \to 0} \frac{2\pi - \sum_i \gamma_i}{\eta^2} = \frac{\sqrt{3}}{2} K_p.$$

C.2.2 Unified Curvature Measures

C.2.2.1 Invariant 2-forms of $\mathbb{E}^3 \times \mathbb{S}^2$

Let $ST(\mathbb{E}^3) = \mathbb{E}^3 \times \mathbb{S}^2$ be the unit tangent bundle of $\mathbb{E}^3$. Thus, elements of $ST(\mathbb{E}^3)$ are pairs (m, ξ), where m is some point of $\mathbb{E}^3$ and ξ is some tangent direction (of unit length) to $\mathbb{E}^3$ at m. Let (m, ξ) be a point of $ST(\mathbb{E}^3)$. Consider an orthonormal frame of $\mathbb{E}^3$ whose third vector is ξ: $(e_1, e_2, e_3 = \xi)$. Then, an orthonormal frame of $T_{(m,\xi)}ST(\mathbb{E}^3)$ is given by the vectors

$$\begin{gathered} \varepsilon_1 = \begin{pmatrix} e_1 \\ 0 \end{pmatrix}, \varepsilon_2 = \begin{pmatrix} e_2 \\ 0 \end{pmatrix}, \varepsilon_3 = \begin{pmatrix} e_3 = \xi \\ 0 \end{pmatrix}, \\ \widetilde{\varepsilon}_1 = \begin{pmatrix} 0 \\ e_1 \end{pmatrix}, \widetilde{\varepsilon}_2 = \begin{pmatrix} 0 \\ e_2 \end{pmatrix}. \end{gathered} \tag{C.2}$$

Using identification of 1-forms with vectors given by the scalar product, we introduce three 2-forms of $ST(\mathbb{E}^3)$ defined at each point (m, ξ), as follows:

$$\omega_A = \varepsilon_1 \wedge \varepsilon_2, \quad \omega_H = \varepsilon_1 \wedge \widetilde{\varepsilon}_2 + \widetilde{\varepsilon}_1 \wedge \varepsilon_2, \quad \omega_K = \widetilde{\varepsilon}_1 \wedge \widetilde{\varepsilon}_2, \tag{C.3}$$

which are invariant under the rigid motion of $\mathbb{E}^3 \times \mathbb{S}^2$.

Compared to the classical surface theory, the restrictions of $\{\varepsilon_1, \varepsilon_2\}$ on the surface are $\{\omega_1, \omega_2\}$ in the structure equation, The restrictions of $\{\widetilde{\varepsilon}_1, \widetilde{\varepsilon}_2\}$ are $\{\omega_{31}, \omega_{32}\}$ respectively. The restrictions of $\{\omega_A, \omega_K, \omega_H\}$ on surfaces are the area element (A.1), Gauss curvature measure (A.2) and mean curvature measure (A.3) straight forwardly.

C.2.2.2 Integral 2-currents of $\mathbb{E}^3 \times \mathbb{S}^2$

By definition, the space of *2-currents* of $ST(\mathbb{E}^3)$ is the topological dual of the space of *2-forms* with compact support on $ST(\mathbb{E}^3)$. *Integral 2-currents* generalize oriented surface patches.

Formally, they are defined as linear combinations of oriented surface patches. The duality $\langle \cdot, \cdot \rangle$ between *2-forms* and *integral 2-currents* is simply given by Lebesgue integration

over each term (that is, each oriented surface): for all *2-forms* ω and *integral 2-currents* $T = \sum_i \lambda_i T_i$, where T_i is represented by an oriented surface, and $\lambda_i \in \mathbb{R}$,

$$\langle T, \omega \rangle = \sum_i \lambda_i \int_{T_i} \omega.$$

C.2.2.3 Normal Cycle of Oriented Surface of $\mathbb{E}^3$

The normal cycle is the generalization of the unit normal bundle.

Définition 54 (**Normal cycle of smooth surface).** *If S is a smooth orientated surface, then its* normal cycle $\mathbf{N}(S)$ *is the current whose support is the unit normal bundle of S with canonical orientation.*

$$\mathbf{N}(S) := \{(p, \mathbf{n}_p) | p \in S, \mathbf{n}_p \in \mathbb{S}^2\}. \tag{C.4}$$

In order to define normal cycles for discrete surfaces, we first define them on simplicial complex, then use the additivity property to define them on the whole mesh.

Définition 55 (**Normal cycle of l-simplex).** *The normal cycle of σ_l is the current (oriented surface) $\mathbf{N}(\sigma_l)$, whose support* spt$(\mathbf{N}(\sigma_l))$ *is the (smooth) hypersurface defined by*

$$\text{spt}(\mathbf{N}(\sigma_l)) = \{(p, \mathbf{n}_p) \in ST(\mathbb{E}^3), s.t. \forall p, q \in \sigma_l, \langle \overrightarrow{pq}, \mathbf{n}_p \rangle \leq 0\} \tag{C.5}$$

and whose orientation is given by the outward normal.

The normal cycle satisfies the following additivity property.

Proposition C.2.4. *Let A and B be currents such that $A \cup B$ and $A \cap B$ are currents as well. Then,*

$$\mathbf{N}(A \cup B) = \mathbf{N}(A) + \mathbf{N}(B) - \mathbf{N}(A \cap B). \tag{C.6}$$

Définition 56 (**Normal cycle of discrete surface).** *If T is a discrete orientated triangulation, let us suppose T can be decomposed into* 0-simplex *,* 1-simplex *, and* 2-simplex *with numbers of m, n, and s respectively. Let σ_l be an elementary l-*simplex*. Then the* normal cycle $\mathbf{N}(T)$ *can be defined as follows:*

$$\mathbf{N}(T) = \mathbf{N}(\{\cup_{i=1}^{m} \sigma_0^i|\} \bigcup \{\cup_{j=1}^{n} \sigma_1^j\} \bigcup \{\cup_{k=1}^{s} \sigma_2^k\}) \tag{C.7}$$

Eqn. (C.7) can be constructed from the normal cycles of simplicies by the additivity property of Eqn. (C.6).

Définition 57 (**The mass of normal cycle).** *The* mass *of the* normal cycle *of a oriented surface S,* $\mathbf{M}(\mathbf{N}(S))$ *is the area of the normal cycle.*

C.2.2.4 Generalized Curvatures

By evaluating the three *invariant 2-forms*: ω_A, ω_H, and ω_K on the *normal cycle* of a oriented surface, we can define the generalized curvatures (i.e. curvature measures) of smooth and discrete surface within a unified framework.

Définition 58 (**Smooth Curvatures Measures).** *Let S be a (oriented) smooth surface of $\mathbb{E}^3$, then the* generalized curvatures *of S restricted on a local region $B \subset S$ are defined as follows:*

$$\begin{aligned}
\mathbf{M}_A^S(B) &= \langle \mathbf{N}(S)|_{B\times\mathbb{S}^2}, \omega_A \rangle = Area(B) \\
\mathbf{M}_H^S(B) &= \langle \mathbf{N}(S)|_{B\times\mathbb{S}^2}, \omega_H \rangle = \int_B H(p)dp \\
\mathbf{M}_K^S(B) &= \langle \mathbf{N}(S)|_{B\times\mathbb{S}^2}, \omega_K \rangle = \int_B K(p)dp
\end{aligned}$$

where $H(p)$ and $K(p)$ denote the mean *and* Gaussian curvatures *of S at point p respectively.*

Définition 59 (**Discrete Curvatures Measures).** *Let T be a (oriented) triangulation mesh of $\mathbb{E}^3$, then the* generalized curvatures *of T restricted on a local region $B \subset T$ are defined as follows:*

$$\begin{aligned}
\mathbf{M}_A^T(B) &= \langle \mathbf{N}(T)|_{B\times\mathbb{S}^2}, \omega_A \rangle = Area(B) \\
\mathbf{M}_H^T(B) &= \langle \mathbf{N}(T)|_{B\times\mathbb{S}^2}, \omega_H \rangle = \sum_{e\in B} H(e) \\
\mathbf{M}_K^T(B) &= \langle \mathbf{N}(T)|_{B\times\mathbb{S}^2}, \omega_K \rangle = \sum_{v\in B} K(v)
\end{aligned}$$

where $K(v)$ denotes the discrete Gauss curvature at the vertex v, $H(e)$ is the discrete mean curvature at the edge e.

C.2.3 Curvature Measure Approximations

Définition 60 (**Projection map).** *Suppose $p \in \mathbb{E}^3$ is a point, the* project point *of p to S is defined as*

$$\mathrm{pr}(p) := \min_{q\in S} d(p,q).$$

It is straightforward to show that $p - \mathrm{pr}(p)$ is along the normal direction at $\mathrm{pr}(p)$.

Some special points may have multiple project points, which form the medial axis of the surface.

Définition 61 (**Medial axis and local feature size).** *Let S be a smooth surface embedded in $\mathbb{E}^3$. The* medial axis *of S is the closure of the set of points with more than one project point on S. The* local feature size $\mathrm{lfs}(p)$ *at point $p \in S$ is the least distance from p to the medial axis. The local feature size of the surface is defined as*

$$\mathrm{lfs}(S) := \min_{p\in S} \mathrm{lfs}(p).$$

The *tubular neighborhood* of the surface S is defined as

$$U_r(S) := \{p \in \mathbb{E}^3 | \min_{q\in S} d(p,q) \le r\}.$$

Définition 62 (**Hausdorff distance).** *Let T be a triangulation and S be a smooth surface embedded in $\mathbb{E}^3$. the* Hausdorff distance *between T and S is defined as*

$$\delta(T,S) = \inf\{r > 0 | T \subset U_r(S), S \subset U_r(T)\}.$$

In particular, given the *projection map* $\mathrm{pr} : T \to S$, let B be a local region of T, $B \subset T$. We denote δ_B as the *Hausdorff distance* between B and $\mathrm{pr}(B)$.

Définition 63 (Closely inscribed). *Let T be a triangular mesh and S be a smooth surface embedded in $\mathbb{E}^3$. T is closely inscribed in S if:*

- *for any vertex $v \in T \Rightarrow v \in S$,*
- *for any vertex $v \in \partial T \Rightarrow v \in \partial S$,*
- *$T \subset U_r(S)$, where $r = \mathrm{lfs}(S)$.*
- *the projection map $\mathrm{pr} : T \to S$ is bijective.*

Définition 64 (Angular deviation). *Let T be a triangulation closely inscribed in a smooth surface embedded in $\mathbb{E}^3$. The* angular deviation *between $p \in T$ and $\mathrm{pr}(p) \in S$ is defined as*

$$\alpha_p = \max_{(p,\mathbf{n}_p)\in \mathrm{spt}\mathbf{N}(T)} \langle \mathbf{n}_p, \mathbf{n}_{\mathrm{pr}(p)} \rangle,$$

where $\mathrm{pr} : T \to S$ is the projection map, $\mathbf{n}_{\mathrm{pr}(p)}$ denotes the unit normal vector of S at point $pr(p)$, $\mathrm{spt}\mathbf{N}(T)$ denotes the support of the normal cycle of T. Furthermore, let B be a local region of T, $B \subset T$, then the angular deviation *between $B \subset T$ and $\mathrm{pr}(B) \subset S$ is defined as*

$$\alpha_B = \sup_{p\in B} \alpha_p.$$

The following theorem gives the approximation error bound for curvature measures in terms of Hausdorff distance δ_B, angular deviation α_B and the mass of normal cycles $\mathbf{M}(\mathbf{N}(T))$ and $\mathbf{M}(\partial\mathbf{N}(T))$ restricted on B.

Theorem C.2.5 (Approximation theorem 71, [Morvan 2008a]). *Let T be a triangulation closely inscribed in a smooth surface S embedded in $\mathbb{E}^3$, let B be a local region of T, $B \subset T$, then the errors of the generalized curvatures between $\mathbf{M}_k^T(B)$ and $\mathbf{M}_k^S(\mathrm{pr}(B))$, $k \in \{A, H, K\}$, are as follows:*

$$|\mathbf{M}_k^T(B) - \mathbf{M}_k^S(\mathrm{pr}(B))| \leq (\delta_B + \alpha_B)\left(\frac{2(1+|h_{\mathrm{pr}(B)}|)}{1-\delta_B|h_{\mathrm{pr}(B)}|}\right)^2 (\mathbf{M}(\mathbf{N}(T)|_{B\times\mathbb{S}^2}) + \mathbf{M}(\partial\mathbf{N}(T)|_{B\times\mathbb{S}^2})),$$

where $|h_{\mathrm{pr}(B)}| = \sup_{\mathrm{pr}(p)\in\mathrm{pr}(B)} |h_{\mathrm{pr}(p)}|$, and $|h_{\mathrm{pr}(p)}|$ denotes the norm of the second fundamental form *of S at $\mathrm{pr}(p)$.*

The following theorem gives the estimation for the masses of normal cycles.

Theorem C.2.6 (Mass of normal cycle [Morvan 2008a]). *With the same conditions as above, one has the bounds for the mass of normal cycles*

1. $\mathbf{M}(\mathbf{N}(T)|_{B\times\mathbb{S}^2}) = O(s(B))$.
2. $\mathbf{M}(\partial\mathbf{N}(T)|_{B\times\mathbb{S}^2}) = O(sd(B))$.

where

$$s(B) = \sum_{t \in T, t \subset \overline{B}} R(t)^2, \tag{C.8}$$

and

$$sd(B) = \sum_{t \in \partial T, t \subset \overline{B}, t \cap \partial B \neq \emptyset} R(t), \tag{C.9}$$

$R(t)$ *is the maximal* circum circle radius *of* T.

Therefore, we need to estimate the bounds for Hausdorff distance, angular deviation and the right hand sides of Eqn.C.8 and C.9. Moreover, we need to show the projection map is bijective.

APPENDIX D

3D Facial Expression Recognition Based on Histograms of Surface Differential Quantities

3D face models accurately capture facial surfaces, making it possible for precise description of facial activities. In this chapter, we present a novel mesh-based method for 3D facial expression recognition using two local shape descriptors. To characterize shape information of the local neighborhood of facial landmarks, we calculate the weighted statistical distributions of surface differential quantities, including histogram of mesh gradient (HoG) and histogram of shape index (HoS). Normal cycle theory based curvature estimation method is employed for the first time on 3D face models along with the common cubic fitting curvature estimation method for the purpose of comparison. Based on the basic fact that different expressions involve different local shape deformations, the SVM classifier with both linear and RBF kernels outperforms the state of the art results on the subset of the BU-3DFE database with the same experimental setting.

D.1 Introduction

Facial expression recognition (FER) has been subject of intensive research effort for the last three decades because of its usefulness in many applications such as human-computer interaction and the analysis of conversation structure [Otsuka *et al.* 2007]. Acknowledging that facial expressions are actuated by contraction of facial muscles, Ekman et al. [Ekman 1972] introduced the Facial Action Coding System (FACS) and evidenced six universal prototypic facial expressions, namely happiness, sadness, anger, fear, surprise and disgust along with neutral. There exists currently an impressive body of results on FER firstly in static 2D images, and then dynamic 2D videos [Fasel & Luettin.J. 2003] and more recently on static 3D scans [L. Yin & Rosato. 2006], [J. Wang & Sun 2006], [Soyel & Demirel. 2007a], [I. Mpiperis & Strintzis 2008], [Tang & Huang 2008a], [B. Gong & Tang 2009], [A. Maalej 2010], [S. Berretti 2010] and dynamic 3D videos [Yin *et al.* 2008a], [Sun & Yin 2008]. With 3D imaging systems readily available, the use of 3D facial data for FER has attracted increasing interest as 3D data are theoretically pose invariant and robust to illumination changes. Furthermore, they also capture accurate geometry information closely sensitive to expression variations.

Existing methods for FER based on static 3D data can be categorized into two streams, i.e. feature based or model based. The first category claims that the distributions of facial surface geometric informations such as gradient and curvature [J. Wang & Sun 2006], distances between pairs of interest landmarks [Soyel & Demirel. 2007a] and local shapes

near landmarks [A. Maalej 2010], [S. Berretti 2010], are closely related to expression categories. These geometric informations are then extracted as features and fed to various classifiers, such as linear discriminant analysis (LDA), support vector machine (SVM) or Neural Network etc., for FER. The main drawback of this kind of approaches is that they require a set of accurately located landmarks. This explains that most of these work made use of the 83 landmarks manually labeled on the BU-3DFE dataset. The second category tries to simulate the physical process of generating expression and explores a generic elastically deformable face model, which can generate universal expressions by adjusting parameters [I. Mpiperis & Strintzis 2008]. In general, this kind of methods needs alignment and normalization steps to find one-to-one correspondence among 3D faces, then the shape deformations between each pair of faces can be represented by model parameters [I. Mpiperis & Strintzis 2008] or feature vectors [B. Gong & Tang 2009] etc, which are further used to perform FER. The downside of these approaches is first their computational cost which is expensive. Furthermore, the fitting process also require existence of some landmarks for initialization and hardly converge in case of opening of the mouth in facial expressions.

In this chapter, we propose a new mesh-based method for FER using two local shape descriptors. To characterize shape information of the local neighborhood of facial landmarks, weighted statistical distributions of surface differential quantities, including histogram of mesh gradient (HoG) and histogram of shape index (HoS) are extracted as local shape descriptors from facial landmarks. These local shape descriptors can be considered as an extension of the popular 2D SIFT feature [D.G.Lowe 2004] (Scale Invariant Feature Transform) to 3D-mesh based discrete surfaces [et al 2010]. Normal cycle theory based curvature estimation method [Cohen-Steiner & Morvan 2003b], [Morvan 2008b], [P. Alliez 2003] is employed for the first time on 3D face models along with the popular cubic fitting curvature estimation method for the purpose of comparison. Based on the basic fact that different expressions involve different local shape deformations, we combine all the local shape descriptors, respectively HoG and HoS, into a global shape descriptor according to a simple early fusion strategy. Finally, all these global features are fed to a SVM classifier using both linear and RBF kernels to perform facial expression recognition. The proposed approach was benchmarked on the BU-3DFE database with the same experimental setting as its main competitors and the experimental results show that the proposed approach outperforms the state of the art.

The remainder of this chapter is organized as follows: differential quantities estimated on mesh-based facial models are introduced in section 2. Section 3 presents the local shape descriptors. Experimental results are discussed in section 4. Section 5 concludes the chapter.

D.2 Estimating Differential Quantities on Triangular Meshes

D.2.1 Estimating Curvature by Normal Cycle Theory based Method

There are many approaches to calculate curvature on triangular meshes based on estimation of curvature tensor. Taubin [Taubin 1995] introduced a 3D curvature tensor from which the principal curvatures directions can be estimated by two of the three eignvectors and the principal curvature can be computed by linear combinations of two of the three

eigenvalues. Cohen-Steiner and Morvan introduced in [Cohen-Steiner & Morvan 2003b], [Morvan 2008b] a discrete definition for the mean, Gaussian curvature and the curvature tensor based on normal cycle theory, and proved that the estimated curvature tensors converge to the true ones of the smooth surface under specific sampling conditions. The basic idea and its discrete form can be carried out as follows [P. Alliez 2003]:

For every edge e of the mesh, there is an obvious minimum (i.e., along the edge) and maximum (i.e., across the edge) curvature. A nature curvature tensor can therefore be defined at each point along an edge, named as generalized curvatures [Morvan 2008b]. This line density of tensors can now be integrated over the arbitrary region B by summing the different contributions from B, leading to the simple expression:

$$\mathscr{T}(v) = \frac{1}{B} \sum_{edges\, e} \beta(e)|e \cap B| \bar{e}\bar{e}^t \tag{D.1}$$

Where v is an arbitrary vertex on the mesh, $|B|$ is the surface area around v over which the tensor is estimated. $\beta(e)$ is the signed angle between the normals to the two oriented triangles incident to edge e (positive if convex, negative if cancave), $|e \cap B|$ is the length of $e \cap B$ (always between 0 and e), and $\bar{e}$ is a unit vector in the same direction as e. In our experiments, we estimate the tensor at every vertex location v, for a neighborhood B of 2-ring. The principal curvatures k_{min} and k_{max} at v can now be estimated by the two maximum eigenvalues of $\mathscr{T}(v)$. Fig. 1 shows the schematic of this method.

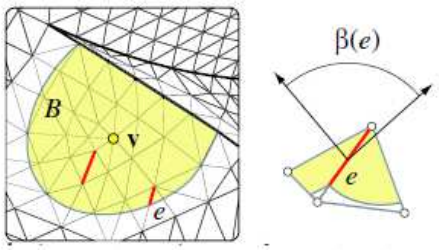


Figure D.1: Illustration of normal cycle theory based curvature estimation method (equation (1) [19]).

Shape index which expresses different shape classes by a single number ranging from 0 to 1 can then be estimated by the following equation:

$$\mathbf{S} = \frac{1}{2} - \frac{1}{\pi} arctan\left(\frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}}\right) \tag{D.2}$$

The first row of Fig.2 shows an example of maximum, minimum curvatures and shape index estimated by this method on a 3D face model.

D.2.2 Estimating Curvature by Local Cubic Fitting based Method

In this paper, we also compare the performance of our shape descriptors using different curvature estimation methods. For this purpose, we also adopt local cubic fitting method [Goldfeather & Interrante 2004a] to estimate curvatures. The basic idea of this method is that: for each vertex p of the 3D mesh, a local coordinate system is defined by taking the

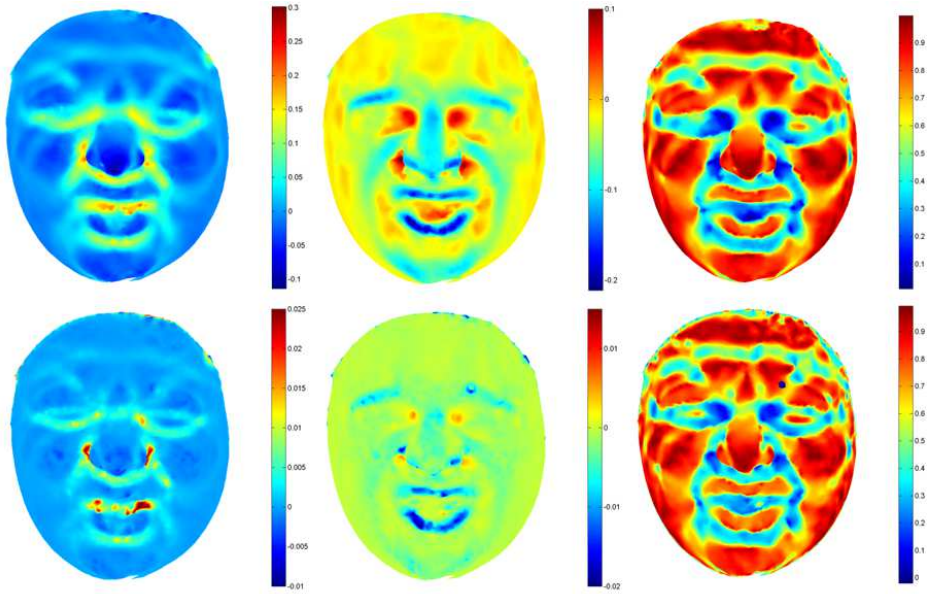


Figure D.2: First row, from left to right: κ_{max}, κ_{min} and shape index estimated by normal cycle theory based method. Second row, from left to right: κ_{max}, κ_{min} and shape index estimated by cubic fitting based method, (model M0044-DI03).

vertex p as an origin and normal vector $n_p = (n_x, n_y, n_z)^T$ as the z axes. Two orthogonal axes, x and y, randomly chosen in the tangent plane perpendicular to the normal vector. The local neighborhood points (2-ring in our paper) and its corresponding normal vectors are first transformed to the local coordinate system, then used for fitting a cubic function and its normal respectively. The cubic function and its normal having the following forms:

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

$$\begin{aligned}(z_x, z_y, -1) = (&Ax + By + 3Dx^2 + 2Exy + Fy^2 + \\ &+ Bx + Cy + Ex^2 + 2Fxy + 3Gy^2, -1)\end{aligned} \tag{D.4}$$

By using least-square fitting method to solve the fitting equations (3) and (4), the Weingarten matrix on a vertex can be computed as:

$$\mathbf{W} = \begin{pmatrix} \frac{\partial^2 z(x,y)}{\partial x^2} & \frac{\partial^2 z(x,y)}{\partial x \partial y} \\ \frac{\partial^2 z(x,y)}{\partial x \partial y} & \frac{\partial^2 z(x,y)}{\partial y^2} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{D.5}$$

The maximum curvature k_{max} and minimum curvature k_{min} then can be estimated as the eigenvaules of the Weingarten matrix. The second row of Fig.2 shows a example of maximum, minimum curvatures and shape index estimated by cubic fitting method on the same 3D face model.

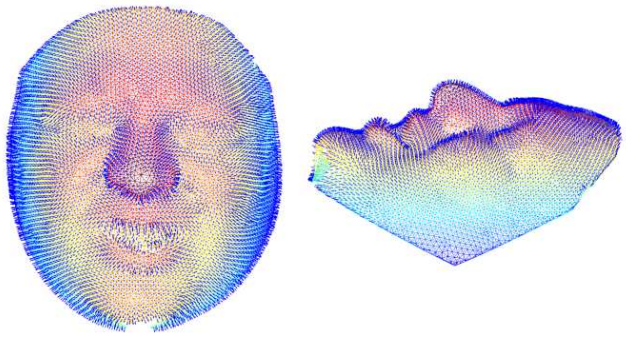


Figure D.3: Illustrated the estimated normal vectors (model M0044-DI03).

D.2.3 Mesh Gradient Estimation

Let the normal vector at p as $n_p = (n_x, n_y, n_z)^T$, which can be estimated by averaging the normal vectors of one-ring faces. According to (3) and (4), the gradient direction and the gradient magnitude can be estimated as follows:

$$\theta = arctan(\frac{n_y}{n_x}) \tag{D.6}$$

$$||\nabla z(x,y)|| = \sqrt{(-\frac{n_x}{n_z})^2 + (-\frac{n_y}{n_z})^2} \tag{D.7}$$

Fig. 3 shows the estimated normal vectors on a 3D mesh-based face model.

D.3 Local Shape Descriptors

D.3.1 Landmarks Selection

A set of landmarks, e.g. eye corners, nose tip, etc. is used for the purpose of FER. For each landmark, two local shape descriptors, namely HoG and HoS, are extracted from its neighborhood. In this work, a neighborhood with a geodesic disk is considered. The experiments were carried out on the BU-3DFE database using the first 60 landmarks defined on regions of eyebrows, eyes, nose and mouth, which are a subset of 83 manually labeled landmarks in the database. The radius of the disk is equal to 22 mm in our experiments. The 60 selected landmarks and an example of local neighborhood of the left mouth corner are shown in Fig. 4.

D.3.2 Local Coordinate System and Orientation Assignment

In practice, we first transform the points within a local neighborhood to a local coordinate system, in which the landmark point is the origin and its normal vector is along the positive z axis. Two perpendicular vectors x and y axis are randomly chosen in the tangent plane. In order to make the descriptor invariant to rotation, each landmark point is assigned one or several canonical orientations according to the dominant direction(s) of gradients in the

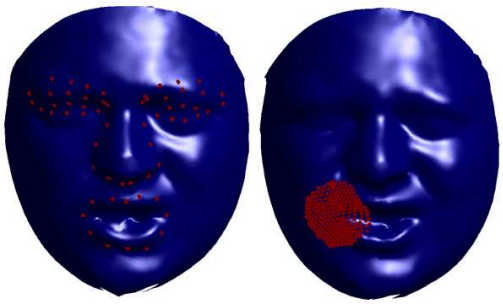


Figure D.4: From left to right, 60 selected manual landmarks, local neighborhood points of the left mouth corner (M0044-DI04 and M0044-DI03).

local tangent plane with 360 bins. Once the canonical orientations are assigned, the local coordinate system rotates in the local tangent plane, making each canonical orientation as new x axis. Now y axis can be computed by cross product of z and x.

In this new local coordinate system, we project all the neighbors of a landmark point to its tangent plane. Eight projected points along to eight quantized directions starting from canonical orientation with a distance of r_1 to the landmark point are fixed. Nine circles centered at the landmark point and its eight neighbors with a radius r_2 can be further located. Fig.5 shows this arrangement.

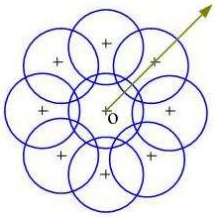


Figure D.5: Canonical orientation (arrow), landmark point (o) and its 8 neighborhood vertices (+) assigned with 9 circles.

D.3.3 Feature Vectors Computed as Histograms of Surface Differential Quantities

In each circle, we calculate the histogram of surface gradient (hog^c) and histogram of shape index (hos^c). For hog^c, we compute the histogram of gradient angle weighted by gradient magnitude. This histogram has 8 bins representing 8 main orientations ranging form 0 to 360 degree. For hos^c, the values of shape index ranging from 0 to 1 are also quantized to 8 bins. Then, all the values of histograms are weighted by a Gaussian kernel with the Euclidian distance to the center point of the circle as the standard deviation. Every histogram with the length of 72 is then normalized and form a feature vector for a single landmark point. Formally, hog^p and hos^p are defined as follows:

$$hog^p = (hog^c_1, hog^c_2, \ldots, hog^c_9) \tag{D.8}$$

$$hos^p = (hos^c_1, hos^c_2, \ldots, hos^c_9) \tag{D.9}$$

The final feature vectors of each face model, respectively HoG and HoS, are then obtained by simply concatenating the corresponding histograms of all 60 manual landmarks in a fixed order, giving birth to a 60 $\times$ 72 = 4320 length bins. They can be represented as follows:

$$HoG = (hog_1^p, hog_2^p, \cdots, hog_{60}^p) \tag{D.10}$$

$$HoS = (hos_1^p, hos_2^p, \cdots, hos_{60}^p) \tag{D.11}$$

In an early fusion strategy, the fusion of these two local shape descriptors, HoG+HoS, are a simple concatenation of these two vectors.

D.4 Experimental Results

The BU-3DFE database [L. Yin & Rosato. 2006] contains 2500 textured 3D face models of 100 subjects with different gender, race, and age. Each subject contains one neutral model and six universal non-neutral expressions: happiness, sadness, anger, fear, surprise and disgust. Also, each non-neutral expression is displayed at four different intensities: low, middle, high and highest. Fig.5 shows some samples of the database with six different expressions of high and highest intensities.

Figure D.6: Examples of six universal expressions, from left to right: anger, disgust, fear, happiness, sadness, surprise. First row, high intensity, Second row, highest intensity.

In our experiments, we used the same set up as in [B. Gong & Tang 2009]. A subset of 60 subjects was randomly selected with two high-intensity models for each of the six facial expressions. In total, 60 $\times$ 12 = 720 3D mesh-face models were selected. Then, 54 and 6 subjects were randomly selected respectively as training set (648 models) and test set (72 models). In practice, since the number of models in test set is very limited (only 12 models for each expression), and people of different race, gender and age may have different facial surface changes when they perform the same expression, the average recognition accuracy obtained by 10 or 20 random experiments varies greatly, from about 50% to more than 90% [B. Gong & Tang 2009] with the same feature set, classifier and parameters setup. To obtain stable average recognition accuracies, we run all of our experiments 1000 times independently. To calculate descriptors, we set r_1 and r_2 equal to 15 mm and 7 mm respectively. HoG and HoS each having 4320 dimensions, it is hard

to have information on their distributions in such high dimension space. We made use of the SVM (LIBSVM [Chang & Lin 2001]) classifier with both liner kernel and RBF (radial basis function) kernel for final FER. The parameter 'gamma' for RBF kernel was set to 0.04 by 8-fold cross-validation. The results are shown in table 1-5. AN, DI, FE, HA, SA, SU designate anger, disgust, fear, happiness, sadness, and surprise, respectively.

Table D.1: Average confusion matrix obtained by HoG

	linear kernel						RBF kernel					
%	AN	DI	FE	HA	SA	SU	AN	DI	FE	HA	SA	SU
AN	**74.0**	8.3	0.8	1.6	15.2	0.1	**66.0**	12.0	3.0	0.9	18.1	0
DI	5.9	**76.5**	10.4	3.7	2.6	0.9	2.3	**80.7**	8.5	3.8	3.2	1.5
FE	5.4	11.7	**63.6**	9.7	5.4	4.2	3.4	8.4	**63.2**	13.2	5.5	6.2
HA	1.1	0.9	16.0	**82.0**	0	0	0.1	1.4	12.6	**85.6**	0	0.3
SA	16.6	2.7	8.2	0	**72.2**	0.3	15.6	4.1	5.7	0	**72.6**	2.0
SU	1.0	2.3	4.9	0.3	1.5	**90.0**	0	2.6	4.1	1.6	0.9	**90.7**
Average	**76.4**						**76.5**					

Table 1 shows the average confusion matrix obtained by the HoG feature using the SVM classifier with linear and RBF kernels. We can find that both linear and RBF kernels achieve relatively high recognition accuracies for happiness and surprise and lower accuracies for other expressions. We observe the same phenomenon already evidenced in other work: anger and fear have comparatively lower classification rates. Anger is confused mainly by sadness while fear is confused mainly by happiness and surprise. Linear kernel works slightly better for anger while slightly worse for disgust and happiness as compared to the RBF kernel. Both of these two kernels achieve almost the same average recognition rate about 76.5% for all six expressions.

Table D.2: Average confusion matrix obtained by normal cycle based HoS

	linear kernel						RBF kernel					
%	AN	DI	FE	HA	SA	SU	AN	DI	FE	HA	SA	SU
AN	**77.0**	8.5	0.9	0	13.3	0.4	**71.5**	10.4	2.2	0	15.2	0.7
DI	7.6	**80.0**	6.2	2.8	3.4	0	3.6	**82.0**	6.5	3.4	4.4	0
FE	5.2	5.9	**70.9**	8.9	5.9	3.2	4.4	8.1	**65.6**	12.0	6.1	4.0
HA	0.4	0.9	4.8	**93.2**	0	0	0.6	1.0	6.5	**91.2**	0	0.7
SA	15.5	1.8	7.3	0	**74.4**	0.9	14.1	2.4	10.1	0	**72.0**	1.5
SU	0.2	1.9	1.9	0.4	0	**95.6**	0.1	1.6	1.4	0.2	0	**96.8**
Average	**81.9**						**79.9**					

Table 2 shows the average confusion matrix obtained by the HoS feature (estimated by normal cycle theory based method) and the SVM classifier of linear and RBF kernels. All the results are better than the accuracies in Table 1 except the one obtained by the RBF kernel for sadness (72%, vs 72.6%). Linear kernel works better than the RBF kernel except disgust (80.0% vs 82.0%) and surprise (95.6% vs 96.8%). The average recognition rates for all six expressions are 81.9% and 79.9% for linear and RBF kernels, respectively.

Table D.3: Average confusion matrix obtained by cubic fitting based HoS

	linear kernel						RBF kernel					
%	AN	DI	FE	HA	SA	SU	AN	DI	FE	HA	SA	SU
AN	**73.1**	7.7	2.7	0.2	16.2	0	**72.3**	9.9	2.6	0	15.2	0
DI	5.1	**80.6**	9.3	3.0	1.7	0.5	3.3	**81.3**	8.5	2.9	2.7	1.2
FE	5.4	4.8	**72.3**	8.7	6.0	2.7	5.5	73.3	**66.6**	9.5	6.7	4.4
HA	0.8	3.9	4.2	**90.5**	0	0.7	0.9	2.2	5.2	**91.1**	0	0.7
SA	16.3	0.6	5.9	0	**76.3**	0.9	12.8	1.5	8.6	0	**76.2**	0.9
SU	0.1	2.3	3.0	0	0.8	**93.9**	0.1	2.3	1.1	0	0.6	**95.9**
Average	**81.1**						**80.6**					

Table 3 shows the average confusion matrix obtained by the HoS feature (estimated by local cubic fitting method) and SVM classifier of linear and RBF kernels. All the results are better than the corresponding results in table 1 except the one obtained by linear kernel for anger (73.1%, vs 74.0%). For the case of linear kernel, the performances are worse than the ones in table 2 except for disgust (80.6% vs 80.0%), fear (72.3% vs 70.9%) and sadness (76.3% vs 74.4%); while for the case of RBF kernel, the performances are a little better than table 2 (80.6% vs 79.9%) especially for sadness (76.2% vs 72%). The two curvature estimation methods achieve comparative results which are in contrast to the conclusion made in [J. Wang & Sun 2006].

Table D.4: Average confusion matrix obtained by HoG+HoS descriptor using a linear kernel with the SVM

	normal cycle based						cubic fitting based					
%	AN	DI	FE	HA	SA	SU	AN	DI	FE	HA	SA	SU
AN	**76.8**	7.6	2.1	0	13.5	0	**76.4**	8.0	1.8	0	13.7	0
DI	7.6	**78.1**	6.6	2.1	5.0	0.7	4.4	**80.2**	10.2	2.8	2.0	0.5
FE	4.6	7.6	**73.2**	7.3	5.1	2.3	5.1	6.2	**73.6**	8.0	5.3	1.7
HA	0.5	0.5	6.8	**91.4**	0	0.8	0.8	2.2	6.5	**90.4**	0	0.8
SA	14.5	1.1	8.3	0	**75.5**	0.6	14.7	0.5	6.2	0	**77.8**	0.8
SU	0	1.7	2.0	0.9	0.8	**94.5**	0	2.0	3.3	0.1	1.0	**93.6**
Average	**81.6**						**82.0**					

Table 4 presents the average confusion matrix obtained by fusing the features HoG and HoS into a single feature according to a simple early fusion scheme, denoted as HoG+HoS (estimated by normal cycle theory and cubic fitting), and a SVM classifier with a linear kernel. Compared to the results in left column of table 2 and the one of table 3 also achived by a SVM with a linear kernel, the fused feature HoG+HoS, when estimated using the normal cycle theory for the computation of curvature, ameliorates the recognition accuracy for sadness and fear while recording a slight performance drop for other expressions. Now when estimated using the cubic fitting technique for the computation of curvature, the fused feature HoG+HoS improves the recognition accuracies for sadness, fear along with anger this time and also records a slight drop for the remaining expressions. Overall, there are

no significant increase of recognition accuracies either using normal cycle or cubic fitting in the case of the early fused feature vector HoG+HoS. This may be caused by the huge dimension (8640×1) of the fused descriptor on the one hand and the different nature of geometric information captured by HoG and HoS. It may be interesting to study some other fusion strategies, for instance late fusion strategy in combining similarity scores.

Table D.5: Comparison of the proposed method with the state of the art [11], [9], [5], [6], [8].

	HoG	HoS	HoG+HoS	Berretti	Gong	Wang	Soyel	Tang
AVE	**76.48%**	**81.86%**	**82.01%**	**77.54%**	**76.22%**	**61.79%**	**67.52**	**74.51%**

Table 5 compares the experimental results achieved by the proposed method (HoG by RBF kernel, HoS by normal cycle and linear kernel, HoG+HoS by cubic fitting and linear kernel) with the ones reported in [B. Gong & Tang 2009] and [S. Berretti 2010]. In fact, the results of the approaches proposed by Gong et al. (Gong) [B. Gong & Tang 2009], Wang et al. (Wang) [J. Wang & Sun 2006], Soyel et al. (Soyel) [Soyel & Demirel. 2007a], and Tang et al. (Tang) [Tang & Huang 2008a], were obtained using the same experimental setting. In Berretti et al. (Berretti) [S. Berretti 2010], 60 subjects were selected randomly from experiment to experiment. It can be seen from the table that the proposed approach using HoG feature achieves comparable result as the others while HoS and HoG+HoS features outperform all the other methods.

D.5 Conclusion and Future Work

In this chapter, we have developed a mesh-based 3D facial expression recognition approach and evaluated it on the BU-3DFE database. The proposed approach is based on two local shape descriptors, namely HoG and HoS, computed from surface differential quantities with the SVM classifier. The surface differential quantities are extracted from the local neighborhoods around a set of landmarks, in this work 60 landmarks from the 83 manually labeled landmarks in the BU3D-FE database. Curvatures are estimated by normal cycle theory based method and cubic fitting method. Both linear and RBF kernels of SVM are employed for classification. The experimental results show that the proposed approach outperforms the state of the art and demonstrate its effectiveness.

In our future work, we will investigate alternative fusion schemes than the simple early fusion scheme used in this work. Furthermore, we also want to study the stability of the proposed approach when landmarks are automatically located for instance by our statistical technique SFAM as proposed in [X. Zhao 2009].

APPENDIX E

3D Facial Expression Recognition via Multiple Kernel Learning of Multi-Scale Local Normal Patterns

In this chapter, we propose a fully automatic 3D person-independent facial expression recognition framework based on Multiple Kernel Learning (MKL) of Multi-Scale Local Normal Patterns (MS-LNPs). To extract discriminative expression features, each aligned 3D facial surface is compactly represented as multiple global histograms of local normal patterns from multiple normal components and multiple binary encoding scales, namely Multi-Scale Local Normal Patterns (MS-LNPs). Expression recognition is finally carried out by modeling multiple kernel learning to efficiently embed and fuse these multiple features. By using the SimpleMKL algorithm with chi-square kernel, we achieved an average recognition rate of 80.14% based on a fair experimental setup. To the best of our knowledge, our result outperforms most of the state-of-the-art results.

E.1 Introduction

Recently, with the tremendous progress of 3D face recognition technology, 3D facial expression recognition has also drawn much attention since the BU-3DFE database [Yin *et al.* 2006b] was made publicly available in 2006 (refer to the recently survey [Fang *et al.* 2011]). However, developing an fully automatic and highly accurate 3D facial expression system is still a huge challenge.

To handle the challenge of accuracy, there are two key issues: facial expression feature and facial expression classifier. Most of the existing feature based methods extract various local or global geometric information as expression features, e.g., curvature based primitive label map [Wang *et al.* 2006b]; distances between facial feature points [Soyel & Demirel 2007b], [Tang & Huang 2008b]; Shape Deformation [Gong *et al.* 2009]; 2.5D SIFT-like descriptor [Berretti *et al.* 2010b]; histograms of surface differential quantities [Li *et al.* 2011d]. Most of these methods reported relative high accuracies while all of them (except [Gong *et al.* 2009]) require a predefined set of facial feature points (i.e., manually labeled landmarks) making them hardly applicable to real-world applications. Meanwhile, except [Berretti *et al.* 2010b] and [Gong *et al.* 2009], all the other methods used an unstable and unfair experimental protocol pointing out by [Gong *et al.* 2009]. In this paper, we address the problem of six prototypical expressions classification by using feature based methods, referring to [Fang *et al.* 2011] for model based methods and other methods for 3D facial action units classification. Once expression features are extracted, many different classifiers can be used,

e.g., Quadratic Discriminant Classifier [Wang *et al.* 2006b]; LDA [Wang *et al.* 2006b], [Tang & Huang 2008b]; Naive Bayes [Wang *et al.* 2006b], [Tang & Huang 2008b]; Nearest Neighbor [Tang & Huang 2008b]; Neural Network [Soyel & Demirel 2007b] and SVM [Wang *et al.* 2006b], [Gong *et al.* 2009], [Berretti *et al.* 2010b], [Li *et al.* 2011d]. In this chapter, we address both of the two key issues. To design a highly discriminative and robust expression feature, three components, in X, Y and Z-plane respectively, of normal vectors estimated from 2.5D range image, are first encoded locally to their corresponding Local Normal Patterns (LNPs) similar to Local Binary Pattern (LBP) works on texture image. Then, to utilize the spatial distribution information of facial normal patterns, each kind of Local Normal Pattern: LNPx, LNPy and LNPz , can be further divided into several patches, from which histograms of LNPs are computed. Finally, the expression feature of a facial surface is represented by a global histogram of LNPs concatenated by facial configuration. Like multi-scale LBP, we encode LNPs with different scales, achieving Multi-Scale Local Normal Patterns (MS-LNPs) [Li *et al.* 2011b]. Based on MS-LNPs, each 3D facial surface now represented by nine feature vectors: coming from three encoding scales and three normal components. How to make best fusion of all these multiple feature vectors to improve the final accuracy becomes a key problem. Fortunately, Multiple Kernel Learning (MKL) based classifier can be a good choice comparing with traditional signal kernel based SVM classifier and recently popular Sparse Representation Classifier (SRC). This is mainly because that MKL can learn different weights for feature vectors extracted from different scales or different normal components, which is usually better than using the same weights by SVM or SRC conducting feature-level or score level fusion. As the generalization of single kernel based classifier, the basic idea of Multiple Kernel Learning (MKL) was first used for genomic data fusion, in which the authors shown that using multiple kernel instead of a single one can enhance the interpretability of the decision function and improve performances [Lanckriet *et al.* 2004].

The chapter is organized as follows: Section 2 introduces the Multi-Scale Local Normal Patterns (MS-LNPs) based expression feature extraction; Section 3 presents the basic idea of Multiple Kernel Learning (MKL) for multi-class expression classification. Experimental results are discussed in section 4. Section 5 concludes the paper.

E.2 Multi-scale local normal patterns

Given a 3D facial surface represented by an $m \times n \times 3$ matrix as follows,

$$\mathbf{P} = \{p_{ij}(x,y,z)\}_{m\times n} = \{p_{ijk}\}_{m\times n\times\{x,y,z\}}, \tag{E.1}$$

where $p_{ij}(x,y,z) = [p_{ijx}, p_{ijy}, p_{ijz}]^T$ represents the 3D coordinates of the point p_{ij}. Its unit normal at each point can be estimated by fitting a local plane. Each of the normal components can be represented by an $m \times n$ matrix:

$$\mathbf{N}(\mathbf{P}) = \begin{cases} \mathbf{N}(\mathbf{X}) = \{n_{ijx}\}_{m\times n}, \\ \mathbf{N}(\mathbf{Y}) = \{n_{ijy}\}_{m\times n}, \\ \mathbf{N}(\mathbf{Z}) = \{n_{ijz}\}_{m\times n}. \end{cases} \tag{E.2}$$

where $-1 \leq n_{ijk} \leq 1$.

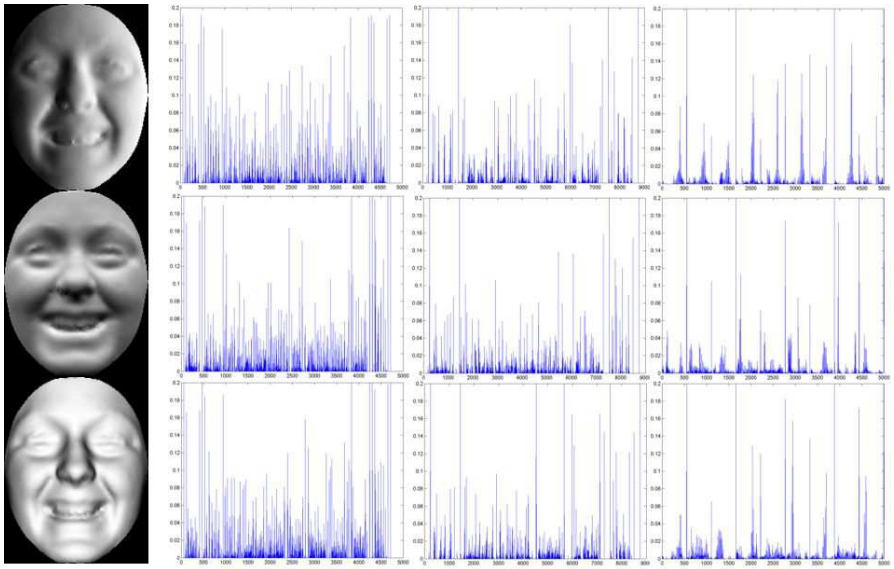


Figure E.1: MS-LNPs: images of normal components X, Y and Z and corresponding features extracted from three scales: $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$ (from left to right).

Inspired by the discriminative power and computational simplicity of LBP describing 2D texture, we encode each normal component as local normal patterns (LNPs). Here we suppose the values of each normal component are just like the intensity values of the nature texture image. Indeed, if we re-scale the normal values from $[-1,1]$ to the range of $[0,255]$, the normal components can be displayed as component-normal images (see Fig.1). Thus, each point of each normal component can be encoded as local binary pattern corresponding to a decimal number. In practice, each facial normal component matrix can be divided into several patches, from which histograms of local normal patterns are extracted respectively, then concatenated by facial configuration to form a global histogram. Finally, the original facial surface is described by nine global histograms extracted from three normal components and three encoding scales (see Fig.1). See [Li *et al.* 2011b] for more details about MS-LNPs.

E.3 Multi-class multiple kernel learning

During last few years, as the generalization of single kernel based support vector machine (SVM), multiple kernel learning (MKL) has proved to be an efficient tool for solving learning problems like classification and regression [Lanckriet *et al.* 2004], especially after Bach et al. proposed the simple-MKL algorithm which enables to effectively tackle large scale problems [Rakotomamonjy *et al.* 2008].

In a binary classification scenario, given the learning set $\{x_i, y_i\}_{i=1}^{M}$, where x_i belongs to some input space $\mathscr{X}$ and y_i is the label of x_i. The MKL makes predictions based a function

of the form

$$f(x) = \sum_{i=1}^{M} \alpha_i^\star \sum_{j=1}^{N} d_j^\star K_j(x, x_i) + b^\star \tag{E.3}$$

where $K_j(\cdot,\cdot), (j = 1,2,\ldots,N)$ is one of the given positive definite basis kernel, $d_j^\star$ is its corresponding weight, and $d_j^\star \geq 0, \sum_{j=1}^{N} d_j^\star = 1$, N is the total number of the kernels, all $\alpha_i^\star$, $d_j^\star$ and $b^\star$ are some coefficients to be learned from the learning set. To solve the MKL problem efficiently, SimpleMKL algorithm involves a weighted l_2 regularization to lead to the following smooth and convex optimization problem (prime MKL):

$$\begin{aligned} \min_{\{f\},b,\xi,d} \quad & \frac{1}{2}\sum_m \frac{1}{d_m}||f_m||^2_{\mathscr{H}_m} + C\sum_i \xi_i \\ s.t. \quad & y_i \sum_m f_m(x_i) + y_i b \geq 1 - \xi_i \quad \forall i \\ & \xi_i \geq 0 \quad \forall i \\ & \sum_m d_m = 1, d_m \geq 0 \quad \forall m \end{aligned} \tag{E.4}$$

where $\mathscr{H}_m = \{f | f \in \mathscr{H}'_m : \frac{||f||_{\mathscr{H}'_m}}{d_m} < \infty\}$, endowed with the inner product $\langle f, g\rangle_{\mathscr{H}_m} = \frac{1}{d_m}\langle f, g\rangle_{\mathscr{H}'_m}$, and $\mathscr{H}'_m$ is a reproducing kernel Hilbert space (RKHS) associated with kernel K_m and inner product $f_m(x) = \langle f(\cdot), K_m(\cdot,\cdot)\rangle_{\mathscr{H}'_m}$.

The dual MKL problem is

$$\begin{aligned} \max_{\alpha_i,\lambda} \quad & \sum_i \alpha_i - \lambda \\ s.t. \quad & \sum_i \alpha_i y_i = 0 \\ & 0 \leq \alpha_1 \leq C \quad \forall i \\ & \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j K_m(x_i, x_j) \leq \lambda \quad \forall m \end{aligned} \tag{E.5}$$

Due to the difficult of optimization the last constraint,

SimpeMKL minimizes a variation version of the prime MKL as follows,

$$\min_d J(d) \quad s.t. \sum_m d_m = 1, d_m \geq 0, \tag{E.6}$$

where

$$J(d) = \begin{cases} \min\limits_{\{f\},b,\xi,d} & \frac{1}{2}\sum\limits_m \frac{1}{d_m}||f_m||^2_{\mathscr{H}_m} + C\sum\limits_i \xi_i \quad \forall i, \\ s.t. & y_i \sum_m f_m(x_i) + y_i b \geq 1 - \xi_i \quad \forall i, \\ & \xi_i \geq 0 \quad \forall i. \end{cases} \tag{E.7}$$

Refer to [Rakotomamonjy *et al.* 2008] for the details of the SimpleMKL algorithm.

Noticing that in our paper, $\{x_i\}_{i=1}^M$ come from multiple sources, i.e., x, y and z components of surface normal vector and three encoding scales of the local normal patterns (LNPs). Features with different normal components or encoding scales usually capture

different and complementary shape information of facial surface, making them have different degree of discriminative power thus should be associate with different weights. And that is exactly what MKL does. Once given kernels associate with different type of features, MKL seeks to the best combination weights of these kernels. One of the most useful kernel used for measurement the similarity of two sets of histogram like feature vectors x and y is chi-square kernel (χ^2 distance)

$$K(x,y) = exp(-\frac{1}{D}\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{\frac{1}{2}(x_i+y_i)}). \tag{E.8}$$

where n is the number of feature vectors, D is the parameter for normalizing the distances. The decision function for binary classification SimpleMKL is

$$D(x) = sign(\sum_{i=1}^{M}\sum_{j=1}^{N}\alpha_i^\star y_i d_j^\star K_j(x,x_i) + b^\star) \tag{E.9}$$

where $\alpha_i^\star$, $d_j^\star$ and $b^\star$ have the same meaning as equation (E.3). Suppose we have a multi-class problem with K class, the multi-class SimpleMKL differ from its binary version by defining a new cost function

$$J(d) = \sum_{p\in\mathscr{P}} J_p(d) \tag{E.10}$$

where $\mathscr{P}$ is the set of all pairs (one-to-one or one-to-rest) to be considered, and $J_p(d)$ is the binary cost function in (E.6).

E.4 Experimental results

We performed six prototypical expressions classification task on the BU-3DFE database [Yin *et al.* 2006b]. A relative fair experimental protocol is used as [Berretti *et al.* 2010b], e.g., during each time of test, first, 60 subjects are randomly selected from 100 subjects (only the two high intensity models of each expressions are selected), then, randomly split 54 to 6 training and testing sets, and repeated 100 times.

Each triangle model is first transferred to one range image stored in a $m \times n \times 3$ coordinate matrix by interpolation technique and then registered by ICP algorithm. As described in section 2, for each training or testing model, we extracted 9 global histograms of local normal patterns: 3 normal components, and 3 encoding scales. The encoding parameters are: $Q(8,1), Q(16,2)$ and $Q(24,3)$ corresponding to local patch sizes of $8\times 8, 20\times 16$ and 40×32, working on the normal component matrices with size of 120×96. (see [Li *et al.* 2011b] for more details.) Three classifiers: score level fusion of sparse representation classifier (SRC) with OMP algorithm and sparse number of 30 [Li *et al.* 2011b]; score level fusion of multi-class SVM classifier [Chang & Lin 2011] and SimpleMKL [Rakotomamonjy *et al.* 2008] with fixed C as 100 are used. To compare the performance of SVM and SimpleMKL, linear kernel, RBF kernel with $\gamma = 3$ (empirical choice) and chi-square kernel (D is set to the average value of the χ^2 similarity matrix of two sets of features) are tested respectively. To simplify test, all the 9 MS-LNPs use the same kind of kernel, SimpleMKL learns the kernel weights associate with different MS-LNPs to achieve the final score.

Table E.1: Average, Minimal and Maximal recognition rates obtained by SRC, SVM and SimpleMKL, kernel I: linear; kernel II: RBF; kernel III: chi-square.

%	Average rate	Min	Max
SRC	**78.36**	59.72	91.67
SVM-kernel I	75.78	56.94	88.89
SVM-kernel II	76.65	63.89	93.06
SVM-kernel III	**78.72**	62.50	95.83
SimpleMKL-kernel I	78.60	65.28	88.89
SimpleMKL-kernel II	78.24	62.50	90.28
SimpleMKL-kernel III	**80.14**	63.89	95.83

Table 1 shows the numerical comparisons. We can see that SimpleMKL performs better than SVM for all the three kernels. SimpleMKL with chi-square kernel achieves best average recognition rate (80.14%), which is better than SRC (78.36%) and SVM-kernel III (78.72%). During the 100 times tests, the differences between maximal and minimal rates of all the methods are around 30% which indicates that the recognition rates are largely depends on the subjects used for training and testing.

Table 2 presents the average confusion matrix achieved by SimpleMKL-kernel III. We find that it performs quite well to classify happiness and surprise, getting the recognition rates of 93.17% and 92.67% respectively. The most likely confused expressions are anger-sadness and fear-disgust-happiness.

Table E.2: Average confusion matrix obtained by SimpleMKL-kernel III

%	AN	DI	FE	HA	SA	SU
AN	**77.92**	6.33	3.08	0.33	11.58	0.75
DI	6.83	**77.17**	7.17	3.58	1.67	3.58
FE	4.42	9.00	**69.25**	9.92	3.67	3.75
HA	0.00	0.25	6.17	**93.17**	0.00	0.42
SA	18.42	2.58	7.08	1.00	**70.67**	0.25
SU	0.08	1.42	4.00	1.00	0.83	**92.67**

Table 3 compares with the performance of SimpleMKL-kernel III and the ones reported in [Gong *et al.* 2009], [Berretti *et al.* 2010b], [Li *et al.* 2011d] and [Lemaire *et al.* 2011]. In fact, Gong et al. [Gong *et al.* 2009] reproduced the approaches of Wang et al. [Wang *et al.* 2006b], Soyel et al. [Soyel & Demirel 2007b], and Tang et al.[Tang & Huang 2008b] under the same experimental setting. We can see that our result is better than others except [Li *et al.* 2011d]. It should be notice that except [Gong *et al.* 2009], [Lemaire *et al.* 2011] and our proposed method, all the others need many manual landmarks. On the other side, choosing fixed 60 subjects in setup I is not as fair as setup II.

E.5 Conclusion

In this paper, person independent facial expression recognition is performed on the whole BU-3DFE database. We propose a novel 3D expression descriptor named multi-scale local

Appendix E. 3D Facial Expression Recognition via Multiple Kernel Learning of Multi-Scale Local Normal Patterns

Table E.3: Comparison of average recognition rates, (**I: Setup [Gong *et al.* 2009]; II: Setup [Berretti *et al.* 2010b]**).

%	I	II
Wang et al. [Wang *et al.* 2006b], [2006]	61.79	-
Soyel et al. [Soyel & Demirel 2007b], [2007]	67.52	-
Tang et al.[Tang & Huang 2008b], [2008]	74.51	-
Gong et al. [Gong *et al.* 2009], [2009]	76.22	-
Berretti et al. [Berretti *et al.* 2010b], [2010]	-	77.54
Li et al. [Li *et al.* 2011d], [2011]	82.01	-
P. Lemaire et al. [Lemaire *et al.* 2011] [2011]	76.22	-
SimpleMKL-kernel III	-	**80.14**

normal patters (MS-LNPs). To efficiently fuse the multiple scales and multiple components of MS-LNPs, multiple kernel learning classifier is employed. Base on a fair setup, we achieved an average recognition rate of 80.14%, which outperforms most of the state-of-the-art results.

APPENDIX F

An Automatic 3D Expression Recognition Framework based on Sparse Representation of Conformal Images

In this chapter, we propose a general and fully automatic framework for 3D facial expression recognition by modeling sparse representation of conformal images. According to Riemann Geometry theory, a 3D facial surface S embedded in $\mathbb{R}^3$, which is a topological disk, can be conformally mapped to a 2D unit disk D through the discrete surface Ricci Flow algorithm. Such a conformal mapping induces a unique and intrinsic surface conformal representation denoted by a pair of functions defined on D, called conformal factor image (CFI) and mean curvature image (MCI). As facial expression features, CFI captures the local area distortion of S induced by the conformal mapping; MCI characterizes the geometry information of S. To model sparse representation of conformal images for expression classification, both CFI and MCI are further normalized by a Möbius transformation. This transformation is defined by the three main facial landmarks (i.e. nose tip, left and right inner eye corners) which can be detected automatically and precisely. Expression recognition is carried out by the minimal sparse expression-class-dependent reconstruction error over the conformal image based expression dictionary. Extensive experimental results on the BU-3DFER dataset demonstrate the effectiveness and generalization of the proposed framework.

F.1 Introduction

F.1.1 Background

Facial expression is one of the most naturally preeminent means for human beings to express their emotions, opinions and intentions. Its automatic recognition has a wide range of applications, including entertainment, psychology studies, smart human-computer interface, etc. [Zeng *et al.* 2009]. Thus, Facial Expression Recognition (FER) has been extensively investigated over the last two decades and mostly focused on 2D facial texture images and videos [Fasel & Luettin 1999], [Samal & Iyengar 1992], [Pantic & Rothkrantz 2000], [Zeng *et al.* 2009].

Despite the great advancements has achieved, 2D/2D video facial texture image based facial expression recognition system still fails to solve the two major challenges: illumination changes and pose variations [Fasel & Luettin 1999], [Samal & Iyengar 1992], [Pantic & Rothkrantz 2000], [Zeng *et al.* 2009]. With the recently fast development of

3D scanning technology, 3D/3D video facial surface based facial expression recognition system becomes the indispensably complementary to the traditional 2D/2D video based methods [Fang *et al.* 2011], [Malassiotis & Tsalakanidou 2011], [Sandbach *et al.* 2012], [Fang *et al.* 2012]. This is mainly because that 3D facial surface can provide accurate geometry and topology structure information in 3D space which is very useful for expression feature extraction. These kinds of expression features are illumination and pose invariant naturally.

F.1.2 Related Work

Different from Action Units (AUs) detection system like [Savran & Sankur 2009], in this paper, we focus on recognizing the six prototypical expressions (i.e. happiness, sadness, anger, fear, surprise and disgust). In this literature, developing a fully automatic 3D landmark detection technique or a fully automatic 3D FER system is still a big challenge. As we have known, most existing 3D FER methods are based on a large set of manual landmarks (e.g. [Wang *et al.* 2006b], [Soyel & Demirel 2007b], [Tang & Huang 2008b], [Tang & Huang 2008c], [Soyel & Demirel 2008], [Berretti *et al.* 2010b], [Maalej *et al.* 2010], [Li *et al.* 2011d], [Sha *et al.* 2011]), which make these kind of methods unrealistic for real world applications.

Only a few efforts have been reported on the automatic recognition of 3D facial expression (e.g. [Venkatesh *et al.* 2009], [Rosato *et al.* 2008], [Mpiperis *et al.* 2008b], [Zhao *et al.* 2010], [Gong *et al.* 2009]). Theses methods can be roughly classified into two categories: i) 2D landmark based [Venkatesh *et al.* 2009], [Rosato *et al.* 2008]; ii) learning based [Zhao *et al.* 2010]. By using the one-to-one correspondence of 2D facial image and 3D facial surface, 2D landmark based methods first automatically detect a set of facial landmarks on the 2D facial image and then map them to its corresponding 3D surface. But automatic 2D facial landmark detection involving illumination changes and pose variations is still very challenging. Meanwhile, learning based methods often train a facial shape/landmark statistical model, which is very computationally expensive and parameter sensitive. However, thanks to the landmark free property of conformal mapping, we only need three landmarks to normalize our conformal images, which makes the normalized conformal images more suitable for modeling sparse representation expression dictionary and classifier. Meanwhile, the selected three main landmarks can be automatically and precisely detected [Szeptycki *et al.* 2009c].

Recently, conformal geometry has been introduced as a powerful surface shape analysis tool, such as surface matching, recognition and stitching [Wang *et al.* 2007] [Gu & Vemuri 2004] surface registration [Zeng *et al.* 2008], 3D face recognition [Szeptycki *et al.* 2010] and 3D FER [Rosato *et al.* 2008] [Savran & Sankur 2009]. Different from the previous works, in this paper, for the first time, we propose to use the unique surface conformal representation, i.e., conformal factor image (CFI) and mean curvature image (MCI), as expression features of the proposed 3D FER framework.

Inspired by [Wright *et al.* 2009b], Ying et al. [Ying *et al.* 2010] proposed to fuse the sparse representation of gray facial images and their local binary patterns (LBP) for 2D FER. Cotter et al. [Cotter 2010] investigated sparse representation for classification of corrupted and occluded 2D facial expressions. Zafeiriou et al. [Zafeiriou & Petrou 2010]

proposed to use sparse representation on the difference facial images for 2D FER. To the best of our knowledge, our work is the first one to solve 3D facial expression recognition problem by modeling sparse representation of well normalized conformal images.

F.1.3 Framework Overview

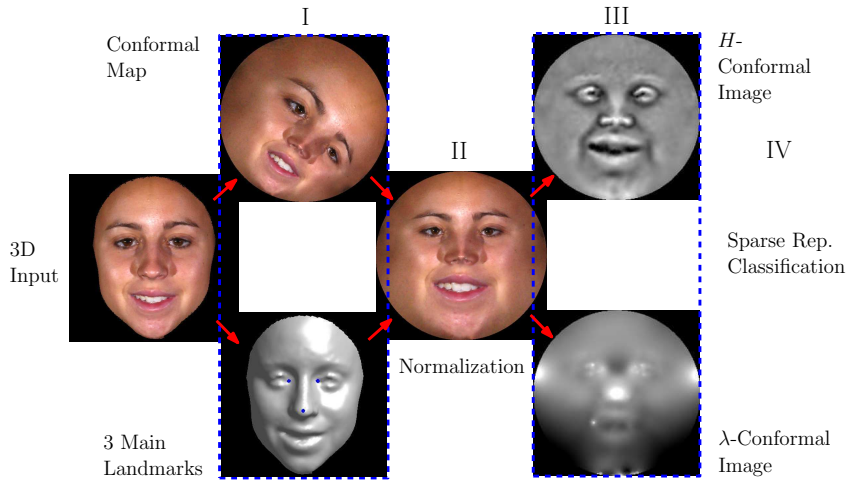


Figure F.1: Overview of the proposed 3D FER framework.

As shown in Fig. 1, our framework consists of four steps: I. Conformal map computation and 3 main landmark detection; II. Conformal map normalization; III. Conformal image based facial surface representation and expression feature extraction; IV. Sparse representation modeling for expression classification. Specifically, given a raw 3D face scan (texture is unnecessary, here, only for better illustration in Fig. 1), we first run the discrete surface Ricci Flow algorithm to compute the conformal map from the original 3D facial surface to the 2D unit disk. Simultaneously, we launch the landmark detection algorithm [Szeptycki *et al.* 2009c] to detect the 3 main landmarks, i.e., the nose tip and the two inner eye corners. To normalize the conformal map, we then apply a Möbius transformation to the conformal map obtained in the first step. The Möbius transformation can be determined by the 3 detected landmarks. Based on the normalized conformal map, the original 3D facial surface can be uniquely represented by a pair of functions: (λ, H), where λ denotes the conformal factor, H denotes the mean curvature. We encode the λ and H information onto the conformal mapping result, then we get two conformal images, such as mean curvature conformal image (MCI) and conformal factor image (CFI). Thus, expression features are extracted as follows: The CFI captures the local area distortion (between the 3D surface and its conformal map in the 2D unit disk) induced by the conformal map. The MCI characterizes the geometry information of the original 3D facial surface. Finally, to build the sparse representation model, expression features (i.e. MCI and CFI) from the training set are arranged according to their expression labels to generate conformal image based expression dictionary. Expression features from the testing set are classified by the minimal sparse expression-class-dependent reconstruction error over the expression dictionary. The main contributions of this paper can be summarized as follows:

- We propose a fully automatic 3D facial expression recognition system. Three main facial landmarks can be automatically and precisely detected and are enough to normalize the conformal map.
- We propose to extract expression features with the unique and complete surface conformal representation, which includes CFI and MCI.
- We propose to solve 3D facial expression recognition problem by modeling sparse representation of well normalized/aligned conformal images.
- We propose a general 3D facial expression recognition framework. Thanks to Riemann geometry theory, any other geometry information (e.g. normal, curvature, shape index, etc.) of the original 3D facial surface can also be mapped to 2D unit disk by computing the conformal map. Meanwhile, any sate-of-the-art 2D texture descriptors (e.g. LBP, HOG, SIFT) can be used on the conformal images to improve their discriminative power.

F.2 Surface conformal mapping

F.2.1 Conformal Mapping and Surface Conformal Representation

Intuitively, *conformal mappings* are angle preserving mappings, which don't distort the angles but areas. Locally, a conformal mapping is a scaling transformation.

Définition 65. *(*Conformal mappings*) Let S_1 and S_2 be two metric surfaces, with Riemannian metrics $\mathbf{g}_1$ and $\mathbf{g}_2$ respectively. A map $\phi : S_1 \to S_2$ is* conformal*, if and only if the pull back metric induced by ϕ satisfies $\phi^* \mathbf{g}_2 = e^{2\lambda}\mathbf{g}_1$.*

The function $\lambda : S_1 \to \mathbb{R}$ represents the area distortion, and is called the *conformal factor*. Specifically, all conformal mappings from the unit disk of a complex plane $\mathbb{C}$, $D = \{|z| < 1 | z \in \mathbb{C}\}$ to itself are the so-called *Möbius transformations*, with the form

$$\phi(z) = e^{i\theta} \frac{z - z_0}{1 - \bar{z}_0 z}, z_0 \in D. \tag{F.1}$$

We can see that Mobius transformation has 3 degree of freedoms: rotation parameter, θ, and translation parameter, z_0. Thus, three points are enough to determine a Mobius transform.

Définition 66. *(*Riemann mapping*) All the simple connected surfaces with a single boundary can be conformally mapped onto the unit disk. Such kind of conformal mapping is called a* Riemann mapping.

Theorem F.2.1 (*Riemann mapping* [Schoen & Yau 1994])**.** *Suppose $(S, \mathbf{g})$ is a topological disk, then there exists a conformal mapping $\phi : S \to D$, which maps S to the unit disk D. All such kind of mappings differ by a Möbius transformation.*

For example, suppose a surface S embedded in $\mathbb{R}^3$, then it has the induced Euclidean metric $\mathbf{g}$. A Riemann mapping $\phi : S \to D$ maps S to the planar disk D, then the original

metric $\mathbf{g}$ can be represented as $\mathbf{g} = e^{2\lambda} dz d\bar{z}$, namely, conformal factor λ determines the first fundamental form of the surface.

Furthermore, assume the mean curvature function of S is $H : S \to \mathbb{R}$, then $H \circ \phi$ is a function defined on the unit disk D, in the following we treat H as a function on D. Under a Riemann mapping, H and λ can determine the second fundamental form of S. The following representation theorem plays a fundamental role in the current work.

Theorem F.2.2 (Surface conformal representation [Gu & Vemuri 2004])**.** *A surface S embedded in $\mathbb{R}^3$ is uniquely determined by the conformal factor and the mean curvature function $\lambda, H : D \to \mathbb{R}$ and its boundary.*

This theorem shows that the pair of conformal factor and mean curvature (λ, H) mainly determines the geometry of the surface. Therefore, we call (λ, H) as *surface conformal representation*.

F.2.2 Surface Ricci Flow

Converting a surface S to its conformal representation (λ, H) is a challenging problem. One of the powerful technical tool for computing conformal mapping is *Surface Ricci Flow*.

Ricci flow was invented by Hamilton for the purpose of proving Poincaré's conjecture [Hamilton 1982]. Ricci flow deforms the Riemannian metric of the surface proportional to its Gaussian curvature, such that the curvature evolves according to a non-linear heat diffusion process, and eventually the curvature becomes constant everywhere. Suppose the Riemannian metric $\mathbf{g} = (g_{ij})$, the Ricci flow is

$$\frac{dg_{ij}}{dt} = -2Kg_{ij}. \tag{F.2}$$

where K is the Gauss curvature of the surface.

Theorem F.2.3 (Ricci Flow)**.** *Surface Ricci flow will converge to constant curvature metric without encountering singularities.*

In practice, for surfaces with boundaries, one can prescribe the target Gaussian curvature for interior points and geodesic curvature on boundary points, the Ricci flow is modified to

$$\frac{dg_{ij}}{dt} = 2(\bar{K} - K)g_{ij}, \tag{F.3}$$

where $\bar{K}$ is the target curvature. Ricci flow still converges and leads to the Riemannian metric with desired curvatures.

F.3 Expression feature extraction: normalized conformal images

We describe the details of expression feature extraction , including conformal mapping computation, normalization and conformal images computation.

F.3.1 Conformal Mapping Computation

We apply discrete surface Ricci Flow algorithm [Zeng *et al.* 2010a] to compute the conformal mapping. First, we choose a point p_0 on the facial surface, and punch a small hole around p_0 to convert the surface to a topological annulus. Then we run Ricci Flow by setting the interior point curvature to be zeros everywhere, and the boundary point geodesic curvature to be zeros as well. The flat metric induced by the Ricci Flow can embed the annulus onto a planar strip isometrically. By applying complex exponential mapping e^z, the strip is mapped to a planar annulus with a concentric inner hole. If the original puncture is tiny, the inner hole will be very small. By filling the inner hole, we obtain a good approximation to the Riemann mapping $\phi : S \to D$.

Due to the Möbius ambiguity, we need to locate main feature points to normalize the conformal mapping. Once conformal mapping is normalized, the conformal images are normalized at the same time.

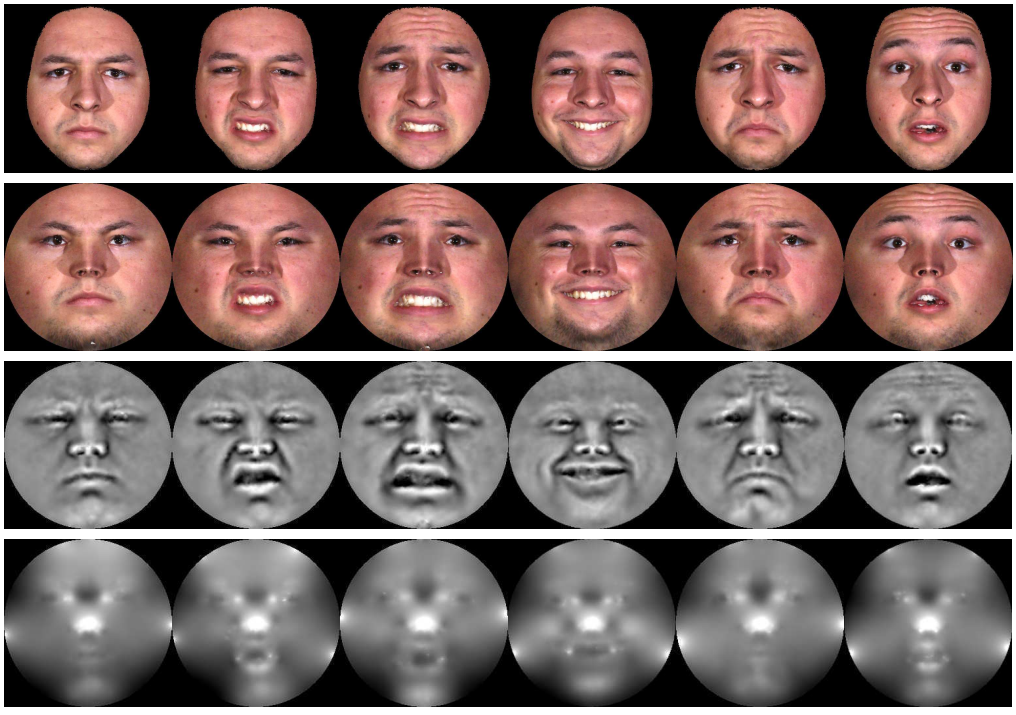


Figure F.2: Conformal images for 3D human facial expressions. The subject is M0021 in BU3D database. 1st row: 6 expression images in camera view; 2nd row: conformal maps with human facial texture; 3rd row: mean curvature images; 4th row: conformal factor images.

F.3.2 Conformal Mapping Normalization

We apply the algorithm in [Szeptycki *et al.* 2009c] to locate the 3 major feature points automatically: the nose tip, the left and right inner eye corners, denoted as $\{p_1, p_2, p_3\}$. By

applying the following Möbius transforamtion

$$\tau(z) = (z - p_1)/(1 - \bar{p}_1 z) \tag{F.4}$$

to move the nose tip to the origin, and a rotation

$$\eta(z) = e^{i\theta} z, \tag{F.5}$$

where $\theta = -arg(\tau(p_2) - \tau(p_3))$ to rotate the line through left and right eye inner corners to be horizontal. Therefore, the normalized Riemann mapping has the form $\eta \circ \tau \circ \phi$. Figure F.2 (row 2) shows the normalized conformal mappings for 6 human faces with different expressions (row 1).

F.3.3 Conformal Images Computation

To achieve the facial surface conformal representation $(\lambda, H) : D \to \mathbb{R}$, the mean curvature and conformal factor of a vertex $v \in S$ are computed as follows: 1) Mean curvature [Morvan 2008a]:

$$H(v) = \frac{\sum_{e \in B} l(e)\beta(e)}{area(B)} \tag{F.6}$$

where v is a vertex of triangular mesh S, B is a local ball neighborhood of a vertex v, $area(B)$ means the area of B, and e is any edge in B, $l(e)$ is the length of e, and $\beta(e)$ is the exterior dihedral angle along the edge. 2) Conformal factor:

$$\lambda(v) = \sum_{T \in B} \lambda(T) = \sum_{T \in B} \frac{\log area(T) - \log \tilde{area}(T)}{2} \tag{F.7}$$

where T is the triangle in B, $area(T)$ is the area on the surface, and $\tilde{area}(T)$ is the area on the unit disk.

CFI and MCI are obtained by scale normalization of λ and H to the range of gray value [0,255]. Fig. F.2 shows the mean curvature images (row 3). and the conformal factor images (row 4). It is obvious that the nose tip area has much larger conformal factors.

F.4 Sparse representation of conformal image for expression classification

F.4.1 Conformal Images based Expression Dictionary

Both the MCI and CFI computed from the training dataset are used for building expression dictionary respectively. For any MCI or CFI with size of $w \times h$, we first reshape it as a m dimensional vector $v \in R_m$ by simply stacking its columns, where $(m = wh)$. Then we arrange the given n_i training samples of the i-th expression class as columns of a matrix $D_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in \mathbb{R}^{m \times n_i}$. Finally, we define a new matrix D for all the expression classes of the training set as the concatenation of the n training samples of all k expression classes:

$$D \doteq [D_1, D_2, \ldots, D_k] = [v_{1,1}, v_{1,2}, \ldots, v_{k,n_k}] \tag{F.8}$$

Specially, since we use the 54 vs. 6 training and testing protocol for six basic expression classification, thus $k = 6$, $n_k = 54 \times 2 = 108$, and $n = 108 \times 6 = 648$. Furthermore, we fixed the ordering of our expression dictionary as anger, disgust, fear, happy, sadness and surprise, corresponding to $D_1, D_2, \ldots, D_6$.

F.4.2 Sparse Representation of Conformal Images

Assume a testing conformal image (MCI or CFI) y is in class i, then the linear representation of y can be approximated by the linear combination of all n training samples,

$$y = Dx + \varepsilon \in \mathbb{R}^m. \tag{F.9}$$

where x is the sparse representation coefficient, whose entries are zeros or closed to zeros except those associated with the i-th class. The basic idea here is that the conformal images of the same facial expression approximately lie in a low dimensional subspace. As pointed out in [Wagner *et al.* 2012], sparse representation model is very sensitive to the face alignment. Thanks to our conformal map normalization step, all the conformal images are well aligned and thus well fitted the sparse representation model.

In order to eliminate the spurious non-zero entries, we seek the sparse solution to Eqn. (F.9) by solving the following optimization problem:

$$\hat{x}_0 = argmin_x ||x||_0 \ \text{ s.t.} ||y - Dx||_2 \leq \varepsilon, \tag{F.10}$$

or its equivalent its l_1 regularization version [Donoho 2006].

$$\hat{x}_1 = argmin_x ||x||_1 \ \text{ s.t.} ||y - Dx||_2 \leq \varepsilon, \tag{F.11}$$

F.4.3 Expression Classification by Class-dependent Reconstruction Error

Similarly to the idea of sparse representation of face recognition in [Wright *et al.* 2009b], we use the expression class-dependent reconstruction error as the rule for expression classification. That is to say, given a test sample y belonging to one kind of prototypical expression, we first estimate its sparse representation $\hat{x}_1$ by optimizing Eqn. (F.11). For each class i, let $\delta_i : R^n \to R^n$ be the characteristic function which selects the coefficients associated with the i-th class. Then, using only the coefficients associated with the i-th class, one can approximate reconstruction the given test sample y as $\hat{y}_i = D\delta_i(\hat{x}_1)$. Finally, we class y by assigning it to the expression class that minimizes the reconstruction error between y and $\hat{y}_i$.

$$\min_i r_i(y) \doteq ||y - D\delta_i(\hat{x}_1)||_2. \tag{F.12}$$

As pointed out in [Wright *et al.* 2009b], sparse representation can be considered as a generalization of nearest subspace (NS) classifier. It considers all the possible supports (with each class and across multiple classes) and chooses the minimal number of the training samples required to represent each test sample.

F.5 Experimental results

F.5.1 Database and Preprocessing

We tested our proposed 3D FER framework on the BU-3DFER dataset [Yin *et al.* 2006b]. It contains 2500 textured 3D facial surfaces of 100 subjects with different attributes such as gender, race, and age. For each subject, it includes one neutral and six prototypical expressions (anger (AN), disgust (DI), fear (FE), happiness (HA), sadness (SA) and surprise (SU)) with 4 intensity levels. All the facial triangular meshes are preprocessed by denoising, hole filling and local re-meshing, to ensure them to be manifolds with disk topology.

F.5.2 Experimental Settings

To set up a fair evaluation for the proposed framework, we used the same experimental setting as [Berretti *et al.* 2010b]. That is to say, only the two high levels of the six prototypical expressions were used for all the experiments (100 subjects, 1200 faces). Meanwhile, for each time, we first randomly selected 60 subjects, and then randomly selected 90% (i.e., 54 subjects, 648 faces) as the training set, 10% (i.e., 6 subjects, 72 faces) as the testing set. To obtain the stable average recognition accuracy, we repeated 100 times. This kind of setting is relatively fair and can avoid the influence caused by the variation of selecting different subjects for training and testing.

The original conformal factor image (CFI) and mean curvature image (MCI) have the same size of 256×256. To achieve a relative low dimension feature vector, we first cropped 20 pixels at each side (216×216) and then resized all the images to be 32×32. Then, each resized image is then reshaped as a vector with dimension of 1024. The Orthogonal Matching Pursuit (OMP) [Pati *et al.* 1993] algorithm with sparse number of 30 was employed to solve Eqn. (F.11) to achieved the sparse representation coefficient $\hat{x}_1$.

F.5.3 Experimental Results

In our test, four experiments were designed for evaluating our framework, including 1) effectiveness and discrimination power analysis, 2) performance evaluations, 3) extensibility analysis, and 4) comparison with the state-of-the-art.

F.5.3.1 Effectiveness and discrimination power analysis

To illustrate the effectiveness and discrimination power of our framework, two simple examples are reported as follows. All the examples are tested on the mean curvature images, conformal factor images are similar.

Example 1. (Right expression classification) For effectiveness analysis, as shown in Fig.F.3, given a testing MCI with disgust expression, we compute its sparse representation coefficients on the MCI based expression dictionary using Eqn. (F.11). We can see that the coefficients in (d) cluster round the range of disgust expression class while sparsely scatter in other classes. Meanwhile, the disgust expression-class-dependent reconstruction error in (e) is the minimal one. Therefore, the testing MCI is rightly classified to disgust expression.

Example 2. (Wrong expression classification) For discrimination power analysis, as shown in Fig.F.4, given a testing MCI with disgust expression, we compute its sparse representation coefficients and find its reconstructed MCI with minimal reconstruction error. We

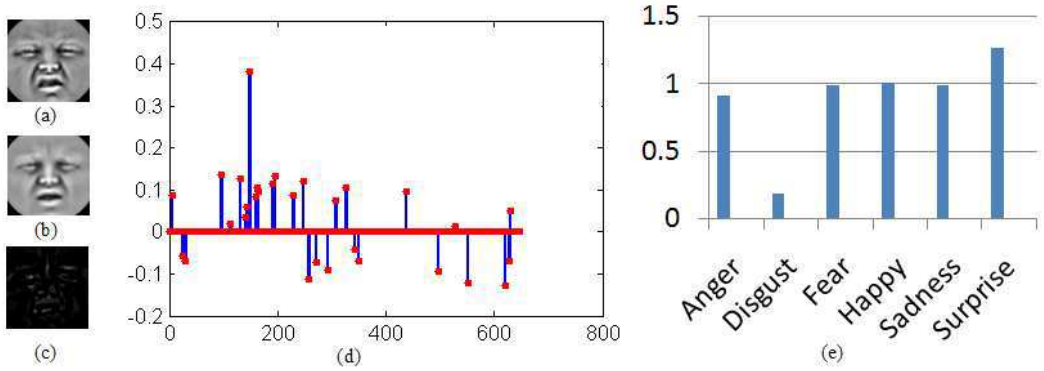


Figure F.3: Example of right expression classification: (a) a testing MCI with disgust expression; (b) the reconstructed MCI with minimal reconstruction error; (c) the minimal reconstruction error image (i.e. the difference of (a) and (b)); (d) the sparse representation coefficients over MCI based expression dictionary (the range of disgust dictionary is from 109 to 216); (e) the expression-class-dependent reconstruction errors.

can see that the distribution of the coefficients in (d) are almost uniform and spread evenly across all the six expression classes. Meanwhile, the expression-class-dependent reconstruction errors are very close to each another except happy and surprise. The minimal one is not disgust but sadness. This example indicates that our framework may have some limitation to discriminate the expressions of disgust, anger, sadness and fear.

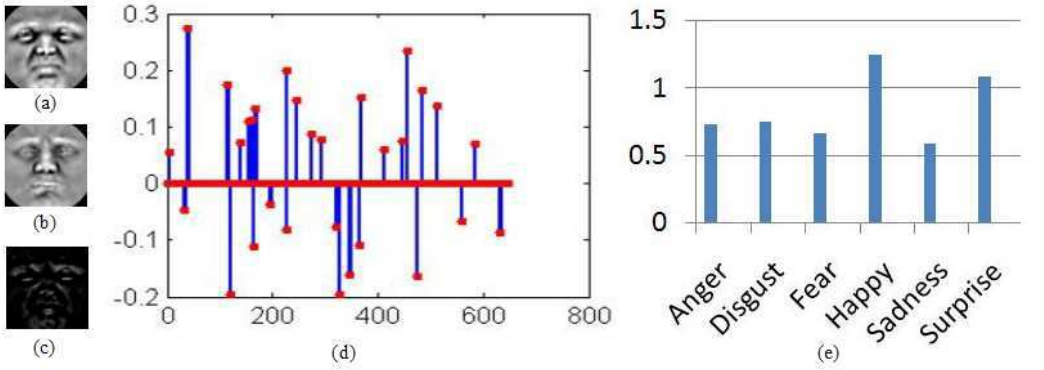


Figure F.4: Example of wrong expression classification: (a) a testing MCI with disgust expression; (b)-(e) have the same meanings as the ones in Fig. F.3.

F.5.3.2 Performance evaluations

To show a visual impression and give a fair evaluation, we demonstrate the performance of one random test based on mean curvature images (MCIs) and the average performance of 100 times tests based on both MCIs and conformal factor images (CFIs).

One random test based on MCI feature: Table F.1 gives the confusion matrix and we

can see that the average recognition accuracy is 79.17%. Specially, all the samples with anger and surprise expressions are rightly classified. And only one sample with happy expression is wrongly classified as sadness. The remaining expressions are confused with the other two or three. Fig. F.5 shows the 72 randomly selected testing samples of MCIs. Fig. F.6 and Fig. F.7 show the corresponding reconstructed MCIs with minimal error and the corresponding error MCIs. The MCIs come from 6 subjects and 6 expression classes (2 intensities for each expression). We can see that even for human beings, recognizing the wrongly classified happy expression sample is also very difficult.

Table F.1: Confusion matrix: one random test based on MCI.

%	AN	DI	FE	HA	SA	SU
AN	**100**	0	0	0	0	0
DI	0	**66.67**	8.33	16.67	0	8.33
FE	25	8.33	**50**	0	16.67	0
HA	0	0	0	**91.67**	8.33	0
SA	25	0	8.33	0	**66.67**	0
SU	0	0	0	0	0	**100**
Average accuracy rate = 79.17%						

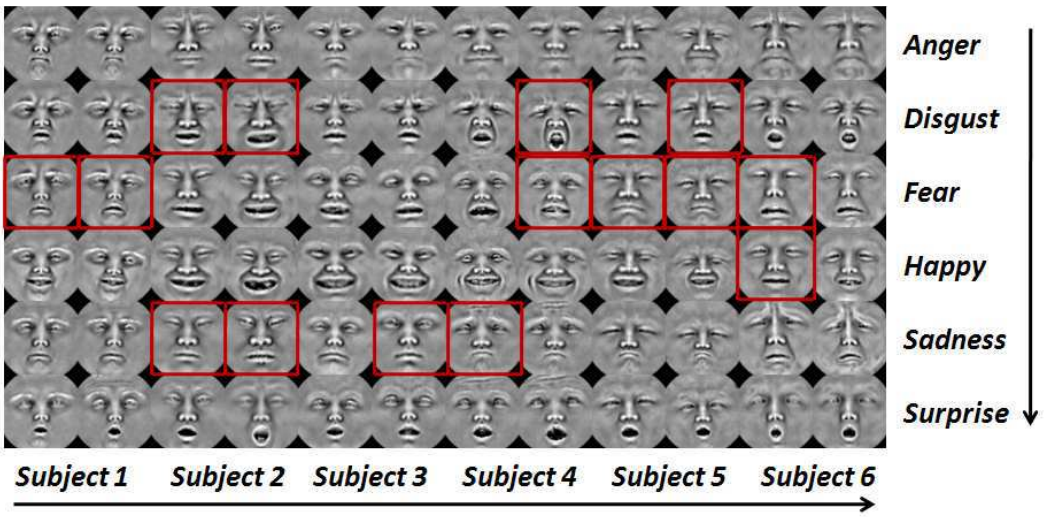


Figure F.5: 72 randomly selected testing MCIs: 6 subjects and 6 classes expressions. The samples of two adjacent columns belongs to the same subject and the samples of each row belongs to the same expression class. The wrong classified samples are indicated by red boxes.

100 times random test based on MCI and CFI features: As shown in Table F.2, the average recognition accuracy based on MCI, CFI and their fusion (on feature level) are 67.28% 61.78% and 68.15%, respectively. We can see that: a) MCI performs better than CFI meaning that MCI contains more geometric information of the original facial surface than CFI. b) The fusion of CFI and MCI performs better than the single one, meaning that they contain complementary geometric information of the original facial surface. These two points

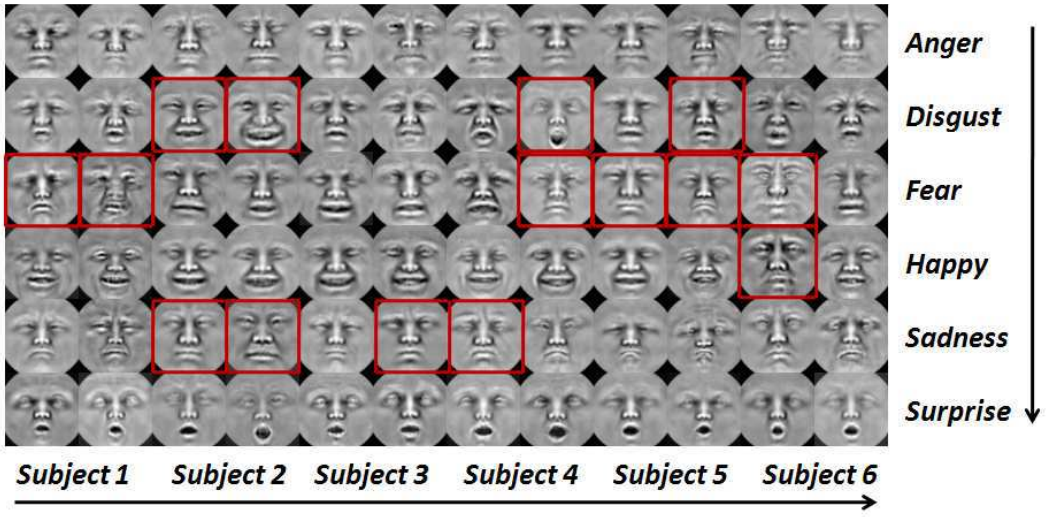


Figure F.6: 72 reconstructed MCIs with the minimal errors corresponding to Fig. F.5. The wrong classified samples are indicated by red boxes.

Figure F.7: The minimal expression-class-dependent reconstruction errors corresponding to Fig. F.5 and F.6.

are consistent to the surface conformal representation theory introduced in section F.2.

Table F.2: Average accuracy: MCI and CFI with SRC classifier.

Features	MCI	CFI	MCI + CFI
Accuracy	67.28%	61.78%	68.15%

Table F.4 shows the average confusion matrix obtained by the feature level fusion of MCI and CFI. The fused feature vector is obtained by simply concatenating of MCI and CFI features with a final dimension of 2048. We can find that the results consistent to the human intuition, e.g. high accuracies are achieved for happiness and surprise, while other ones get relatively lower recognition rates. Further more, anger and sadness greatly confuse with each other, and disgust and fear are confused by all of the six expressions.

Table F.3: Comparison of 3D FER frameworks and performance on the BU-3DFER dataset (L-Automatic: the framework needs leaning or training a model; F-Automatic: the framework is fully automatic and learning free.)

Methods	**Landmarks**	**Automaticity**	**Expression Feature**	**Classifier**	**Performance**		
					I	**II**	**III**
Wang et al. [Wang *et al.* 2006b] [2006]	64 Manual	Non-Automatic	Primitive label map	LDA	83.60%	61.79%	-
Soyel et al. [Soyel & Demirel 2007b] [2007]	11 Manual	Non-Automatic	Distance of landmarks	Neural network	91.30%	67.52%	-
Soyel et al. [Soyel & Demirel 2008] [2008]	83 Manual	Non-Automatic	Distance of landmarks	Neural network	93.72%	-	-
Tang et al. [Tang & Huang 2008b] [2008]	83 Manual	Non-Automatic	Distance of landmarks	AdaBoost + LDA	95.10%	74.51%	-
Tang et al. [Tang & Huang 2008c] [2008]	83 Manual	Non-Automatic	Distance and slopes	SVM	87.10%	-	-
Berretti et al. [Berretti *et al.* 2010b] [2010]	27 Manual	Non-Automatic	2.5D SIFT	SVM	-	-	77.54%
Berretti et al. [Berretti *et al.* 2011a] [2011]	9 Automatic	F-Automatic	2.5D SIFT	SVM	-	-	78.43%
Li et.al [Li *et al.* 2011d] [2011]	60 Manual	Non-Automatic	3D HoG+HoS	SVM		82.01%	-
Gong et al. [Gong *et al.* 2009] [2009]	Not used	L-Automatic	Depth changes	SVM	-	68.77% 76.22%	-
Zhao et.al [Zhao *et al.* 2010] [2010]	19 Automatic	L-Automatic	Texture and geometry	BBN	82.30%	-	-
Mpiperis et.al [Mpiperis *et al.* 2008b] [2008]	Automatic	L-Automatic	Bilinear model parameters	Max-Likelihood	90.50%	-	-
Lemaire et.al [Lemaire *et al.* 2011] [2011]	21 Automatic	L-Automatic	ICP-distance	SVM	-	75.76%	-
Our method 1	3 Automatic	F-Automatic	MCI + CFI	SRC	-	-	**68.15%**
Our method 2	3 Automatic	F-Automatic	LBP-(MCI + CFI)	SRC	-	-	**70.93%**

F.5.3.3 Extensibility analysis

In this section, we want to point out that our proposed framework is very general and has strong extensibility. Different from the raw MCI and CFI, it would be more discriminative if we continue to extract some compact and histogram based 2D texture descriptors (e.g. LBP, HOG, SIFT) on these conformal images. In addition, except MCI and CFI, we can also use other conformal images, such as geometry map and normal map. Furthermore, we can also extract more discriminative 2D local texture descriptors (e.g. LBP, HOG, SIFT) on them.

In this paper, we demonstrate a simple experiment to show how to improve the discriminative of raw conformal images by using LBP descriptor. We continue to extract LBP descriptor on MCIs and CFIs respectively and then use feature level fusion of them. Here, we only use the single scale LBP operator with the parameters $P = 8$ and $R = 1$, i.e. $LBP_{8,1}$. The same experimental setting of 100 times random test is used. The average accuracy is 70.93 % shown in Table F.3. We can see that the performance of LBP-(MCI+CFI) is better than MCI+CFI. This experiment indicates that extracting descriptor based on conformal images is quite promising to improve the performance of our proposed framework.

F.5.3.4 Comparison with the state-of-the-art

Table F.3 demonstrates the comparison of 3D FER frameworks and their performances on the BU-3DFER dataset. To give a comprehensive evaluation, the following elements are considered: number of landmarks, automaticity of landmark detection, automaticity of 3D FER system, expression feature, classifier, experimental setting and performance. Three experimental settings are considered: I, II and III were first used in [Wang *et al.* 2006b], [Gong *et al.* 2009] and [Berretti *et al.* 2010b], respectively. In setting I, 60 subjects are selected and fixed, and then 54 training and 6 testing (54 vs. 6) protocol is used, the average accuracy is obtained as the mean accuracies of 20 times tests. In setting II, the authors also selected and fixed 60 subjects as [Wang *et al.* 2006b], and used 54 vs. 6 protocol. But the average accuracy is obtained as the mean accuracies of 1000 times tests. The drawback of setting I and II is that the ID of selected 60 subjects are not listed for fair comparison. Meanwhile, the results may be largely different if different groups of 60 subjects are used. Compared with setting I and II, III is more fair, it randomly selects 60 subjects from 100 subjects, and then runs the 54 vs. 6 protocol, the average accuracy is achieved as the mean accuracies of 100 times tests. From Table F.3, we can get the following conclusions:

- Most of the existing 3D FER frameworks need a large set of manual landmarks, therefore are non-automatic [Wang *et al.* 2006b], [Soyel & Demirel 2007b], [Soyel & Demirel 2008], [Tang & Huang 2008b], [Tang & Huang 2008c], [Berretti *et al.* 2010b], [Li *et al.* 2011d]. Only a few methods can detect a large set of landmarks automatically but by a learning way [Zhao *et al.* 2010], [Lemaire *et al.* 2011], and some others are landmark-independent [Gong *et al.* 2009], [Mpiperis *et al.* 2008b] but need to train a generic model; They are learning based automatic frameworks. As an advantage, our framework only needs three landmarks, which can be learning free and automatically detected, therefore, is fully automatic.

- The proposed expression feature has deeply theoretical rigorous and depends on the surface conformal representation theorem. Meanwhile, the discriminative power of our feature can be improved by local texture descriptor like LBP.

- Different from the classifiers used in other methods, we present to build sparse representation model over the well normalized conformal images for expression classification.

Table F.4: Average confusion matrix: MCI+CFI

%	AN	DI	FE	HA	SA	SU
AN	**59.58**	10.91	5.58	0.83	22.91	0.17
DI	9.92	**63.67**	12.83	3.67	3.83	6.83
FE	7.58	11.08	**51.00**	15.25	7.91	7.16
HA	1.33	3.08	9.58	**84.58**	0.33	1.08
SA	27.50	2.83	6.83	1.58	**59.58**	1.67
SU	0.91	2.91	3.75	1.42	0.50	**90.50**
Average accuracy = 68.15%						

- For performance evaluation, the work [Gong *et al.* 2009] gives the reproduced results of [Wang *et al.* 2006b], [Soyel & Demirel 2007b] and [Tang & Huang 2008b] in setting II and points out that all the results are largely dropped from setting I to setting II. The result of [Gong *et al.* 2009] changes from 68.77% to 76.22% when the learned generic neutral face model was used. Some other model-based methods, such as [Zhao *et al.* 2010], [Mpiperis *et al.* 2008b] and [Lemaire *et al.* 2011], need an offline training phase, unsuitable to use setting III. As a comparison, the framework of [Berretti *et al.* 2010b] has higher accuracy than ours in setting III but it uses 27 manual landmarks and 85 uniform sample landmarks near nose, mouth and face contour areas, therefore is highly non-automatic. It should notice that the improvement version of [Berretti *et al.* 2010b] in [Berretti *et al.* 2011a] can automatically detect 9 special keypoints.

F.6 Conclusion

In this paper, we propose a fully automatic and general person-independent 3D facial expression recognition framework based on sparse representation of conformal images. By computing conformal mapping, the original 3D facial surface can be conformally mapped to a 2D unit disk. Meanwhile, three automatically detected landmarks are used for conformal map normalization. Based on the normalized conformal map, the original surface can be uniquely represented by two functions defined on 2D unit disk, mean curvature and conformal factor. This two functions correspond to two conformal images, CFI and MCI, which are used for expression feature extraction. Based on these two well normalized conformal images, we design sparse representation model for expression classification. In our experiments, we give the fair evaluation results and show the effectiveness and extensibility analysis of our framework. Finally, we illustrate a comprehensive comparison with the state-of-the-art 3D FER frameworks on the BU-3DFER dataset.

APPENDIX G

A Group of Facial Normal Descriptors for Recognizing 3D Identical Twins

In this chapter, to characterize and distinguish identical twins, three popular texture descriptors: i.e. local binary patterns (LBPs), gabor filters (GFs) and local gabor binary patterns (LGBPs) are employed to encode the normal components (x, y and z) of the 3D facial surfaces of identical twins respectively. A group of facial normal descriptors are thus achieved, including Normal Local Binary Patterns descriptor (N-LBPs), Normal Gabor Filters descriptor (N-GFs) and Normal Local Gabor Binary Patterns descriptor (N-LGBPs). All these normal encoding based descriptors are further fed into sparse representation classifier (SRC) for identification. Experimental results on the 3D TEC database demonstrate that these proposed normal encoding based descriptors are very discriminative and efficient, achieving comparable performance to the best of state-of-the-art algorithms.

G.1 Introduction

As the performance of 2D still-image face recognition system in constrained environments continues to increase [Grother *et al.*] and the high enough accuracies achieved by existing 3D face recognition algorithms against public datasets like FRGC v2.0, focus is shifting from methods that improve face recognition performance in general, to the ones that handle specific failure cases, caused by interference factors such as illumination conditions, pose, lapse of time, facial expression changes [Klare *et al.* 2011]. Recently, the new scenario of distinguishing identical twins becomes a challenging problem of traditional face recognition system since the quite strong similarity of their 2D facial appearance and 3D facial geometric shapes. Some recent studies have evidenced these challenges [Sun *et al.* 2010], [Phillips *et al.*], [Klare *et al.* 2011], [Vijayan *et al.* 2011b], [Vijayan *et al.* 2011a].

To the best of our knowledge, Sun *et al.* [Sun *et al.* 2010] were the first to evaluate the face recognition performance on a dataset consisting of multiple biometric traits (fingerprint, face, and iris) of identical twins. They tested face modality using the Cognitec FaceVACS system on 134 subjects (66 pairs of twins and two sets of triplets). They revealed that their face system could distinguish non-twins better than identical twins. Meanwhile, they concluded that the distribution of identical twin impostor was more similar to the genuine distribution than the general impostor distribution.

Phillips *et al.* [Phillips *et al.*] investigated the performance of three of the top submissions to the Multiple Biometric Evaluation (MBE) 2010 Still Face Track [Grother *et al.*] on a dataset of twins acquired at Twins Day [Twi] in 2009 and 2010. Their experimental

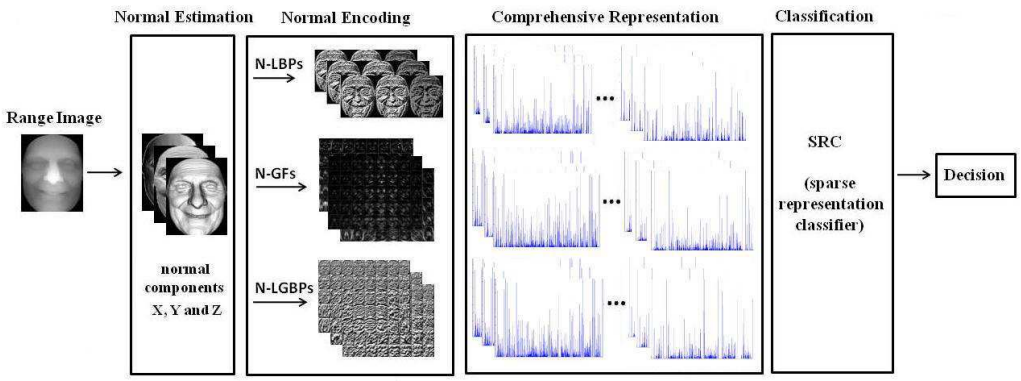


Figure G.1: Framework of our proposed method.

dataset is composed of images taken from 126 pairs of identical twins (252 people) collected on the same day and 24 pairs of identical twins (48 people) with images collected one year apart. They revealed that the results were largely dropped by the variations of lighting conditions (studio and outside); expressions (neutral and smiling); gender and age.

Klare *et al.* [Klare *et al.* 2011] proposed to classify facial features into three levels, i.e. appearance features (Level 1), local features (Level 2) as well as detailed features (Level 3), and studied the feature distinctiveness of Level 2 (MLBP and SIFT) and Level 3 (facial marks) with respect to distinguish identical twins. They indicate that these features which perform well in twin identification are not always consistent with the ones good at recognizing non-twin faces.

Later, Vijayan *et al.* [Vijayan *et al.* 2011b] evaluated the performance of four state-of-the-art 3D face recognition algorithms against the largest dataset of 3D facial scans of twins, namely the 3D Twins Expression Challenge (3D TEC) dataset which contains 107 pairs of identical twins. They found that some algorithms perform very well on the FRGC v2.0 dataset but vastly degrades in performance on the 3D TEC dataset, especially distinguishing the cases combining factors related to facial similarity and the variation of facial expression. Their results show that 3D face recognition of identical twins in the presence of varying facial expression is far from a solved problem, but good performance is possible [Vijayan *et al.* 2011b]. In another work of Vijayan *et al.* [Vijayan *et al.* 2011a], they also pointed out that distinguishing between identical twins with the variation of facial expression is a very challenging problem.

Recently, Li *et al.* [Li *et al.* 2011b] proposed a 3D face descriptor namely Multi-Scale Local Normal Patterns (MS-LNPs) which encodes the three normal components of facial surface using Multi-scale LBP operators. The recognition was carried out by the weighted sparse representation of the encoded facial normal descriptor (MS-LNPs) to enhance the robustness of their system to resist the variation of facial expression. The weights of each facial physical component were learned from training datasets. They achieved a rank-one recognition rate of 96.3% which is comparable to the state-of-the-art tasks. One of the important conclusions they draw in their work lies in that: the encoded normal information is much more discriminative than both the original normal information and the encoded geometric information (i.e., depth images or range images). On the other side, as we know, except Local Binary Patterns (LBP) [Ahonen *et al.* 2004b], Gabor Filters (GFs)

[Wiskott *et al.* 1997b], and Local Gabor Binary Patterns (LGBPs) [Zhang *et al.* 2005] are also two popular face descriptors successfully used in the state-of-the-art 2D face biometric systems. Later, LBP [Huang *et al.* 2011c] and GFs [Cook *et al.* 2006], [Xu *et al.* 2009] have also been exploited to describe 3D facial shape and applied to 3D face recognition systems. In contrast to what we do in this study, all these 3D face descriptors directly encode the geometric information of 3D facial surfaces.

Inspired by the two facts mentioned above, in this paper, we propose two new ways to encode normal information, including multi-scale and multi-orientation Gabor Filters (GFs) and Local Gabor Binary Patterns (LGBPs). Together with MS-LNPs [Li *et al.* 2011b], we achieve a group of facial normal descriptors consisting of Normal Local Binary Patterns descriptor (N-LBPs, i.e. MS-LNPs), Normal Gabor Filters descriptor (N-GFs) and Normal Local Gabor Binary Patterns descriptor (N-LGBPs). The effectiveness of the proposed descriptors are conducted on the 3D TEC dataset, and we find that all these descriptors are very discriminative to distinguish identical twins.

The rest of the chapter is organized as follows. The framework overview of the proposed method is presented in section 2. Section 3 introduces the method of facial normal estimation. Facial normal encoding and representations are explained in section 4. Section 5 describes the sparse representation classifier. In section 6, we show the experimental and algorithmic settings as well as the results. Section 7 concludes the chapter.

G.2 Framework Overview

As shown in Fig. G.1, our framework consists of four steps: 1) normal estimation; 2) normal encoding; 3) comprehensive representation and 4) classification. Specifically, given a raw 3D facial scan, we first launch the preprocessing step to normalize the range image to an $m \times n \times 3$ matrix (i.e., x, y and z coordinates). Based on the range image, we estimate its three normal components (x, y, and z) by local plane fitting method (see Sec. 3). For each normal component, we propose to use three kinds of encoding and representation methods and thus we achieve a group of normal descriptors: N-LBPs, N-GFs and N-LGBPs. Since all the three descriptors comprehensively encode normal information either by multiple scales or orientations, we make use of two manners to fuse them for classification. Feature-level fusion which concatenates all the features extracted at different scales and orientations; and score-level fusion which creates one feature vector at one scale or orientation for the classifier to calculate an individual similarity score, and all these scores are further fused for the following step. Considering its powerful classification ability, in this study, we apply sparse representation classifier (SRC) for the classification. Due to the high dimension of the feature vector of N-GFs and N-LGBPs, we employ PCA for dimensionality reduction. The final similarity measurement of each descriptor computed by combining the ones of three normal components are used for decision making.

G.3 Facial Normal Estimation

Given a range image based face model represented by an $m \times n \times 3$ matrix as follows,

$$\mathbf{P} = [p_{ij}(x,y,z)]_{m\times n} = [p_{ijk}]_{m\times n\times\{x,y,z\}}, \tag{G.1}$$

where $p_{ij}(x,y,z) = (p_{ijx}, p_{ijy}, p_{ijz})^T, (1 \le i \le m, 1 \le j \le n, i, j \in \mathbb{Z})$ represents the 3D coordinates of the point p_{ij}. Let its unit normal vector matrix $(m \times n \times 3)$ be

$$\mathbf{N}(\mathbf{P}) = [n(p_{ij}(x,y,z))]_{m \times n} = [n_{ijk}]_{m \times n \times \{x,y,z\}}, \tag{G.2}$$

where $n(p_{ij}(x,y,z)) = (n_{ijx}, n_{ijy}, n_{ijz})^T, (1 \le i \le m, 1 \le j \le n, i, j \in \mathbb{Z})$ denotes the unit normal vector of p_{ij}. We have $\|n(p_{ij}(x,y,z))\|_2 = 1$. As described in [Hoffman & Jain 1987], the normal vector **N(P)** of range image **P** can be estimated by using local plane fitting method. That is to say, for each point $p_{ij} \in \mathbf{P}$, its normal vector $n(p_{ij})$ can be estimated as the normal vector of the following local fitted plane:

$$S_{ij} : n_{ijx}q_{ijx} + n_{ijy}q_{ijy} + n_{ijz}q_{ijz} = d, \tag{G.3}$$

where $(q_{ijx}, q_{ijy}, q_{ijz})^T$ represents any point within the local neighborhood (5×5 window in our paper) of point p_{ij} and $d = n_{ijx}p_{ijx} + n_{ijy}p_{ijy} + n_{ijz}p_{ijz}$.

Figure G.2 shows one example of three normal component matrices (images) estimated from the original range image. we can see that normal components highlight the informative clues of the raw range image, e.g. the details in the forehead region, which corresponds to the conclusion in [Li *et al.* 2011b] . To simplify, each normal component in equation (2) can be represented by an $m \times n$ matrix:

$$\mathbf{N}(\mathbf{P}) = \begin{cases} \mathbf{N}(\mathbf{X}) = [n_{ij}^x]_{m \times n}, \\ \mathbf{N}(\mathbf{Y}) = [n_{ij}^y]_{m \times n}, \\ \mathbf{N}(\mathbf{Z}) = [n_{ij}^z]_{m \times n}. \end{cases} \tag{G.4}$$

where $-1 \le n_{ij}^k \le 1, k \in \{x,y,z\}$.

Figure G.2: From left to right: the original range image, and its normal images of component x, y and z.

G.4 Facial Normal Encoding and Representation

Thanks to the matrix form of these normal components in equation (G.4), we can encode and characterize each of them using the similar way of feature extraction as for 2D texture images. In this paper, we propose to use LBP, Gabor and LGBP descriptors to encode the discriminative information from each of these normal components.

G.4.1 Normal Local Binary Patterns

Our Normal Local Binary Patterns (N-LBPs) is directly inherited from MS-LNPs proposed in [Li *et al.* 2011b], where three scales of LBP, i.e., $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$, ($Q_{n,m}$ denotes a neighborhood of m sampling points on a circle of radius of n) operator were used. Formally, given a point p_{ij}, its normal component noted as $n_{ijk}(0)$, the derived LNPs decimal value is:

$$LNPs(Q_{n,m}(p_{ij})) = \sum_{q=1}^{m-1} t(n_{ijk}(q) - n_{ijk}(0))2^q, \tag{G.5}$$

where $t(x) = 1, \text{if } x \geq 0; t(x) = 0, \text{if } x < 0$. $LNPs(Q_{n,m})$ encodes local normal variations of each normal component as decimal value, noted by $e(\{n_{ijk}\}_{m\times n}), k \in \{x,y,z\}$. See Fig. G.3 for an example of $LNPs(Q_{1,8})$ representation of three normal components.

To describe a local shape region, the normal component is encoded as the histogram of LNPs as follows:

$$H = \sum_{i,j} I\{e(\{n_{ijk}\}_{m\times n}) = r\}, r = 0,\ldots,R-1, \tag{G.6}$$

where R is the encoded decimal number, for $Q_{1,8}, R = 2^8 = 256$. $I\{A\} = 1, \text{if } A \text{ is ture, else } I\{A\} = 0$. This histogram contains the local micro-patterns of normal component over the whole face model. To utilize spatial information of facial shape, each facial normal component can be further divided into several patches, from which local normal patterns histograms H are extracted; then concatenated by facial configuration to form a global histogram G. Please see [Li *et al.* 2011b] for more details.

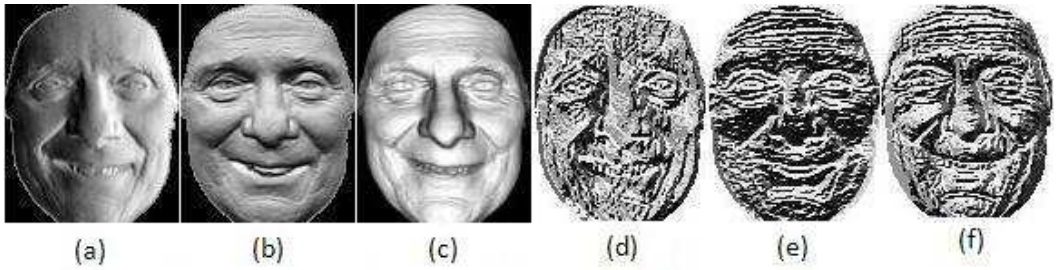


Figure G.3: (a) to (c): normal images of component x, y and z; (d) to (e), the corresponding N-LBPs (LNPs) representation $Q_{1,8}$.

G.4.2 Normal Gabor Filters

Inspired by the motivation that the smooth properties of depth Gabor image cannot describe the facial features in detail [Xu *et al.* 2009], our proposed Normal Gabor Filters descriptor (N-GFs) characterizes facial surfaces by Gabor Filters on its normal components. The mathematical expression of 2D Gabor wavelets can be defined as follows [Lee 1996]:

$$\psi(z) = \frac{k_{\mu,v}^2}{\sigma^2}\exp(-\frac{k_{\mu,v}^2 z^2}{2\sigma^2})[\exp(ik_{\mu,v}z) - \exp(-\frac{\sigma^2}{2})], \tag{G.7}$$

where $z = (x,y)$, and μ and v define the orientation and scale of the Gabor wavelets. $k_{\mu,v}$ is define as follows:

$$k_{\mu,v} = k_v e^{i\Phi_\mu}, \tag{G.8}$$

where $k_v = k_{max}/f^v$ and $\Phi_\mu = \pi\mu/8$. k_{max} is the maximum frequency, and f is the spacing factor between Gabor filters in the frequency domain. In this study, we use Gabor filters with five different scales, $v \in \{0,\ldots,4\}$, and eight orientations, $\mu \in \{0,\ldots,7\}$, with the parameters $\sigma = 2\pi, k_{max} = \pi/2$ and $f = \sqrt{2}, asin$[Xu *et al.* 2009].

By the convolution operation of Gabor Filters and normal components, we achieve the Gabor based representation of normal components:

$$O^k_{\mu,v}(x,y) = [n^k_{ij}(x,y)]_{m\times n} * \psi(x,y), k \in \{x,y,z\}, \tag{G.9}$$

where * denotes the convolution operator.

Given the normal component k, orientation μ and scale v, at each point (x,y), $O^k_{\mu,v}(x,y)$ is a complex number. We compute its magnitude. Thus, totally, for each normal component k, we can achieve 40 (5×8) Gabor magnitude images. Please see Fig. G.4 for an example of normal component z. The comprehensive description of Normal Gabor Filters (N-GFs) can be obtained by feature level fusion or score-level fusion.

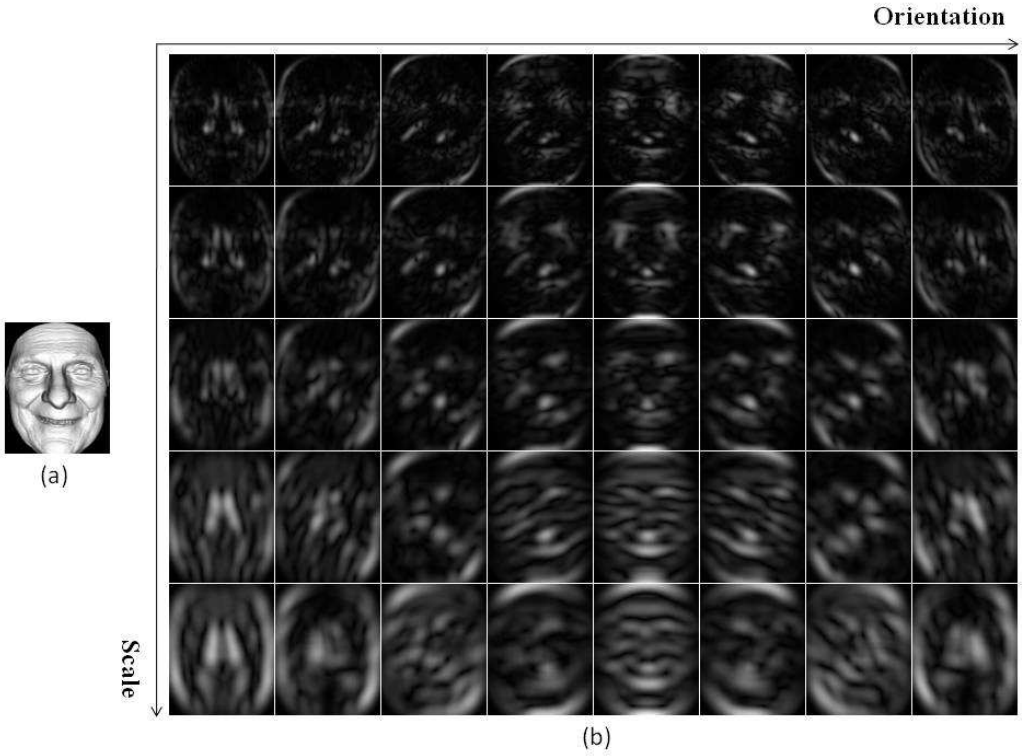


Figure G.4: (a). Normal component z. (b) Its Gabor representation.

G.4.3 Normal Local Gabor Binary Patterns

It have been proved that the combination of Gabor filters and LBP (i.e. Local Gabor Binary Patterns, namely LGBP for short) can improve the performance of LBP for 2D face recognition [Zhang *et al.* 2005]. Analogously, to characterize 3D facial surfaces, our proposed Normal Local Gabor Binary Patterns descriptor (N-LGBPs) employ LGBP operator

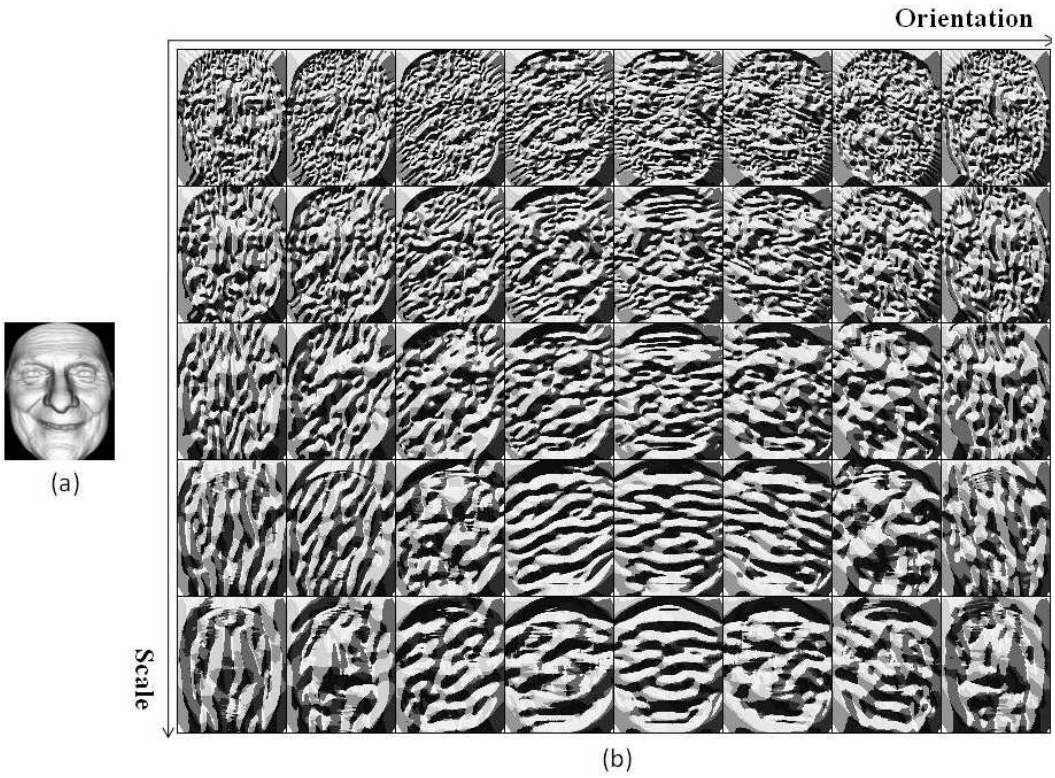


Figure G.5: (a). Normal component z. (b) Its LGBP representation.

to encode the facial normal components. That is to say, we continue to encode the micro-patterns of normal Gabor magnitude images by the LBP operator.

Formally, given a point p_{ij}^k of normal component k, the derived N-LGBPs decimal value is:

$$N\text{-}LGBPs_{\mu,v}^k(Q_{n,m}(p_{ij}^k)) = \sum_{q=1}^{m-1} t(O_{\mu,v}^k(q) - O_{\mu,v}^k(0))2^q, \tag{G.10}$$

where $Q_{n,m}$, t, q, 0 have the same meaning as Eq. (G.5), and $O_{\mu,v}^k$ represents the Gabor representations of normal component k as Eq. (G.9). See Fig. G.5 for an example of $N\text{-}LGBPs_{\mu,v}^z(Q_{1,8})$ based normal representation. Like LBP, to avoid the loss of spatial information of facial normal representation by histograms, each Normal Gabor Magnitude image is further divided into several non-overlapping regions. The global representation can be obtained by concatenating (by facial configuration) all the N-LGBPs histograms extracted from each sub-regions. Similarly, the comprehensive description of Normal Local Gabor Binary Patterns (N-LGBPs) can be achieved using feature level fusion or score-level fusion.

G.5 Sparse Representation Classifier

Based on the face subspace model: a well-aligned frontal face image under different lighting conditions and various facial expressions lies close to a special low-dimensional lin-

ear subspace spanned by sufficient training samples from the same subject, Wright *et al.* [Wright *et al.* 2009c] modeled the face recognition problem by solving the l_1 minimization sparse representation. They also proposed sparse representation classifier (SRC) for robust 2D face recognition.

Analogously, we assume that a well-aligned frontal test 3D face model represented by a feature vector under different facial expressions approximately lies in a linear subspace spanned by the 3D faces in the training set (represented by the same type of facial features) associated with the same subject.

That is, given n_i training samples of i-th subject, $A_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in \mathbb{R}^{m \times n_i}$, any test sample $y \in \mathbb{R}^m$ from the same subject can be represented by:

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \ldots + \alpha_{i,n_i} v_{i,n_i}, \tag{G.11}$$

where $\alpha_{i,j} \in \mathbb{R}, j = 1, 2, \ldots, n_i$.

However, the case is not exactly the same, in our experiments, the training set is composed of one face model from each subject (gallery set) ($n_i = 1$). This problem of insufficient training samples introduces a new model error, noted by $\varepsilon \in \mathbb{R}^m$. Model (8) can be modified as:

$$y \approx \alpha_i v_i = \alpha_i v_i + \varepsilon, \tag{G.12}$$

where $y \in \mathbb{R}^m$, $v_i \in \mathbb{R}^m$ and $\alpha_i \in \mathbb{R}$ represent a probe face, a gallery face from the same subject and their linear scalar factor respectively.

Considering the whole gallery set with n 3D faces, each of which belongs to one subject, $A \doteq [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{m \times n}$ and any probe $y \in \mathbb{R}^m$, (G.12) can be rewritten as

$$y = A\mathbf{x} + \tilde{\varepsilon}, \tag{G.13}$$

where $\mathbf{x} = [0, \ldots, 0, \alpha_i, 0, \ldots, 0]^T \in \mathbb{R}^n$ is the coefficient vector whose entries are zero except the one associated with the i-subject. Sparse coefficients $\mathbf{x}$ in (G.13) can be solved as the following l_1 minimization problem:

$$\hat{x}_1 = argmin_x ||x||_1 \text{s.t.} ||A\mathbf{x} - y||_2 \leq ||\tilde{\varepsilon}||_2, \tag{G.14}$$

We employ OMP [Pati *et al.* 1993] algorithm to solve (G.14) and compute the residuals:

$$r_i(y) = ||y - A\delta_i(\hat{x}_1)||_2, i = 1, 2, \ldots, n. \tag{G.15}$$

where δ_i is a characteristic function which selects coefficient associated with the i-th gallery. Finally, the index of minimal $r_i(y)$ corresponding to the identity of y.

G.6 Experimental and Algorithmic Settings

G.6.1 Datasets and Preprocessing

To evaluate the performance of our proposed descriptors to distinguish identical twins, as used in [Vijayan *et al.* 2011b], all our results were reported on the 3D TEC dataset. 3D TEC is a subset of the Twins Days 2010 dataset acquired at the Twins Days Festival in Twinsburg, Ohio [Twi]. The whole dataset contains 266 subject sessions, with the 3D scans in the dataset containing two scans: one with a neutral expression and another with

a smiling expression. There are 106 sets of identical twins, one set of triplets, and the rest are non-twins. The subset (i.e. 3D TEC) only includes the 3D face scans acquired in the first session and thus consists of of 107 pairs of twins (two of the triplets are included as the 107th set of twins). More instructions of data parameters can be found in [Vijayan *et al.* 2011b].

All 3D face models were preprocessed using the tool developed by Szeptycki *et al.* [Szeptycki *et al.* 2009b], including spike removal, hole filling. Then manually labeled nose tips provided by the dataset were used for face cropping and an ICP fine registration. After normal estimation from each normalized range image, three normal component matrices: $[n_{ij}^x]_{m\times n}, [n_{ij}^y]_{m\times n}$ and $[n_{ij}^z]_{m\times n}$ are resized to the same size of 120×96 for the following encoding step.

G.6.2 Experimental Settings

We did experiments according to the same experimental protocol as defined in [Vijayan *et al.* 2011b]. To be specific, one person in each pair of twins was arbitrarily labeled as Twin A and the other as Twin B. Verification and identification experiments were performed using four different gallery and probe sets as shown in Table 1.

No.	Gallery	Probe
I	A Smile, B Smile	A Neutral, B Neutral
II	A Neutral, B Neutral	A Smile, B Smile
III	A Smile, B Neutral	A Neutral, B Smile
IV	A Neutral, B Smile	A Smile, B Neutral

Table G.1: Gallery and probe sets for cases I, II, III, and IV. "A Smile, B Neutral" means that the set contains all images with Twin A smiling and Twin B neutral [Vijayan *et al.* 2011b].

In Case I, all the images in the gallery set possess a smiling expression while all the images in the probe set have a neutral expression. Case II reverses these roles of Case I. Both the two cases model a scenario that the gallery has one expression and the probe has another. As stated in [Vijayan *et al.* 2011b], in the verification scenario, both the match and non-match pairs of gallery and probe images have different expressions and in the identification scenario, theoretically the main challenge would be to distinguish between the probe image and the image of his/her twin in the gallery since they look similar.

In Case III, Twin A smiling and Twin B neutral make up of the gallery set; while Twin A neutral and Twin B smiling as the probe compose the probe set. Case IV reverses these roles of Case III. These two cases model a worst scenario in which the system does not control for the expressions of the subject in a gallery set of twins. As pointed out in [Vijayan *et al.* 2011b] as well, in the verification scenario, the match pairs would have opposite expressions like in Cases I and II but the non-match pairs that are of the same pair of twins would have the same expression and in the identification scenario, theoretically the main challenge would be to distinguish between the probe image and the image of his/her twin in the gallery. This is more difficult than Cases I and II since the expression of the probe face is different from his/her image in the gallery but is the same as the image of his/her twin in the gallery.

G.6.3 Algorithmic Settings

On Normal Local Binary Patterns descriptor (N-LBPs, i.e. MS-LNPs), as introduced in [Li *et al.* 2011b], each normal component matrix is empirically divided into local patches of sizes 12×12, 20×16 and 40×32 for the operators LNPs of $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$ respectively. Thus three LBP histograms are achieved with dimensions of 4,720, 8,748 and 4,995. Each of this histogram can be used for score level fusion based comprehensive representation directly. For feature level fusion based comprehensive representation, the three histograms are concatenated to achieve the final feature vector (with a dimension of 18,463).

On Normal Gabor Filters descriptor (N-GFs), for feature level fusion based comprehensive representation, each Gabor magnitude image (size of 120×96) is first down-sampled by a factor of 8 and then reshaped to a column vector with a dimension of 180 (15×12). The final feature vector (with a dimension of 7,200) is achieved by concatenating all the 40 column vectors. For score level fusion based comprehensive representation, each Gabor magnitude image is directly reshaped to a column vector with dimension of 11,520 (120×96).

On Normal Local Gabor Binary Patterns descriptor (N-LGBPs), empirically, we use LBP operator of $Q_{2,16}$ with local patches of 20×16. For feature level fusion based comprehensive representation, all the Normal Gabor Magnitude images are first down-sampled by a factor of 2 before extracting LBP histograms. By concatenating all the 40 LBP based histograms, the final feature thus has a dimension of 87,480 ($40 \times 2,187$). For score level fusion based comprehensive representation, each Normal Gabor Magnitude image is directly encoded by the LBP histogram with a dimension of 8,748.

PCA is employed to reduce the huge dimensions of the N-GFs and N-LGBPs descriptors. For all the cases, including the two descriptors with two kinds of comprehensive representations, according to the number of the gallery samples, we fix the compacted dimension of features at 205. To solve (G.14), OMP algorithm with sparse coding number of 30 was used.

G.6.4 Experimental Results

G.6.4.1 Comparison of comprehensive representations

To evaluate the influences of different comprehensive representations related to different facial normal descriptors, we compared their rank-one performance at both feature level and score level fusion according to Case I. Since the effectiveness of fusion different normal components have been proved by Li *et al.* [Li *et al.* 2011b], all the results shown this study are the final one obtained by fusing the three normal components. From Tab. G.2, we can see that score level fusion is better for N-LBPs; while feature level fusion is better for N-GFs and N-LGBPs. Therefore, we report our following results based on their better schemes, i.e. score level fusion for N-LBP, feature level fusion for N-LGFs and N-LGBPs.

Appendix G. A Group of Facial Normal Descriptors for Recognizing 3D Identical Twins

Descriptor	Rank-1 Recognition Rate	
	score-level fusion	feature-level fusion
N-LBPs	**94.86%**	92.52%
N-GFs	88.79%	**93.93%**
N-LGBPs	93.46%	**96.73%**

Table G.2: Comparison of feature level and score level fusion based comprehensive representation (tested in Case I))

G.6.4.2 Comparison of identification results

Tab. G.3 shows the rank-one recognition rate of our proposed method (noted as Alg. 5) and the algorithms reported in [Vijayan *et al.* 2011b]. We can find that N-LGBPs performs best among the group of proposed normal descriptors for all the four cases. N-LBPs and N-GFs achieved comparable results for all the four cases. All the proposed descriptors work better for Cases I and II than Cases III and IV. Compared with other algorithms, N-LGBPs displays comparable results to the best results (given by Alg. 4) for all these four cases. This indicates that our proposed normal descriptors (especially N-LGBPs) prove very high discrimination to recognize identical twins.

Algorithm	Rank-1 Recognition Rate			
	I	II	III	IV
Alg.1 (E_{pkn}) [Faltemier *et al.* 2008b]	93.5%	93.0%	72.0%	72.4%
Alg.1 (E_{minmax}) [Faltemier *et al.* 2008b]	94.4%	93.5%	72.4%	72.9%
Alg.2 (SI) [Huang *et al.* 2011f]	92.1%	93.0%	83.2%	83.2%
Alg.2 (eLBP) [Huang *et al.* 2011c]	91.1%	93.5%	77.1%	78.5%
Alg.2 (Range PFI) [Huang *et al.* 2011d]	91.6%	93.9%	68.7%	71.0%
Alg.2 (Text. PFI) [Huang *et al.* 2011d]	95.8%	96.3%	91.6%	92.1%
Alg.3 [Gokberk *et al.* 2006]	62.6%	63.6%	54.2%	59.4%
Alg.4 [Kakadiaris *et al.* 2007b]	**98.1%**	**98.1%**	**91.6%**	**93.5%**
Alg.5 (N-LBPs)	94.9%	96.3%	89.3%	88.3%
Alg.5 (N-GFs)	93.9%	94.4%	89.3%	90.2%
Alg.5 (N-LGBPs)	**96.7%**	**96.7%**	**92.5%**	**93.5%**

Table G.3: Comparison of rank-one scores of our method (Alg. 5) and the state-of-the-art methods.

G.6.4.3 Comparison of verification results

In [Wright *et al.* 2009c], they tested the outlier rejection (i.e. face validation) capability of sparse representation classifier using the False Accept Rate (FAR) and the Verification Rate (VR) curve (FAR-VR curve, i.e. Receiver Operating Characteristic (ROC) curve). Our verification scenario is like the case of subsection IV-A.2 in [Wright *et al.* 2009c], where the outlier is the same as the imposter.

As [Vijayan *et al.* 2011b], we evaluate the verification results using True Accept Rate at 0.1% False Accept Rate (TAR at 0.1% FAR) and Equal Error Rate (EER). Figures G.6, G.7 and G.8 show the Receiver Operating Characteristic (ROC) curves of the verification experiments for N-LBPs, N-GFs and N-LGBPs respectively. From the figures, we can

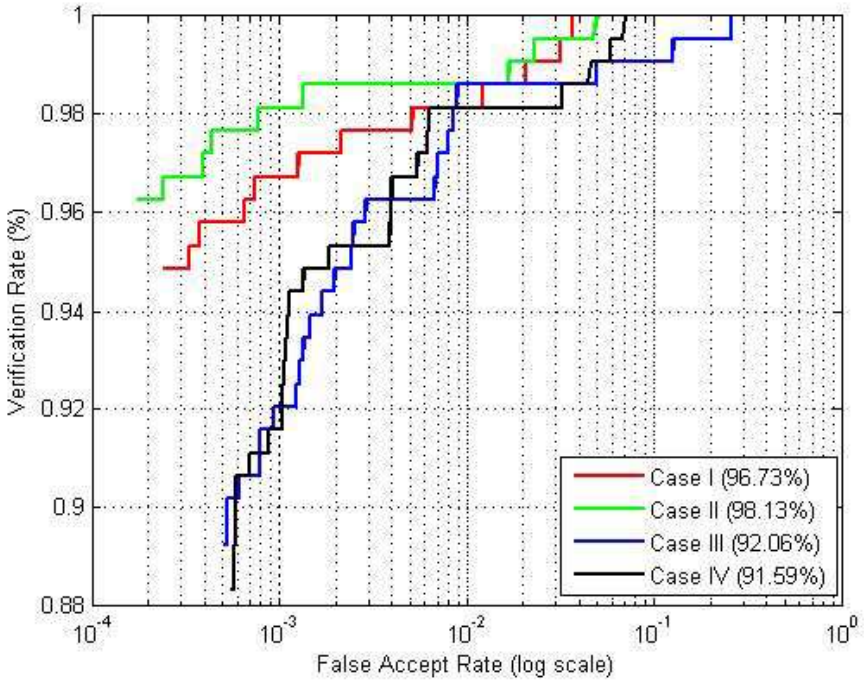


Figure G.6: ROC curves in four cases of our method based on N-LBPs. The legend shows TAR at 0.1% FAR.

find out that the TAR (at 0.1% FAR) results of Case I and Case II are obviously better than that of Case III and Case IV for all the three descriptors. Tab. G.4 and G.5 show the comparison of TAR and EER of our method (Alg. 5) and and the tasks reported in [Vijayan *et al.* 2011b]. We can see that for TAR, N-LGBPs performs better than N-LBPs and N-GFs (except N-LBPs in Case II). N-LBPs achieves the best performance as Alg. 4 in Case II. For other three Cases, the results achieved by N-LGBPs are slightly below but still comparable to the best ones in the literature. For EER, all the three descriptors achieve comparable accuracies to the best so far reported in this database.

G.7 Conclude and Perspective

In this chapter, we proposed a group of facial normal descriptors consisting of Local Normal Patterns (N-LBPs, i.e. MS-LNPs), Normal Gabor Filters (N-GFs) and Normal Local Gabor Binary Patterns (N-LGBPs). Their discriminations for distinguishing 3D identical twins were evaluated in both identification and verification scenarios. Meanwhile, both feature level fusion and score level fusion based comprehensive representations of each descriptor were compared. We achieved the state-of-the-art results.

In our future work, we will mainly investigate the follow issues: 1) As used in [Li *et al.* 2011b], we will study the learning-based weighted sparse representation method for N-GFs and N-LGBPs to resist facial expression variations. 2) Except PCA, other more efficient techniques for dimensionality reduction and feature selection methods, such as

Appendix G. A Group of Facial Normal Descriptors for Recognizing 3D Identical Twins

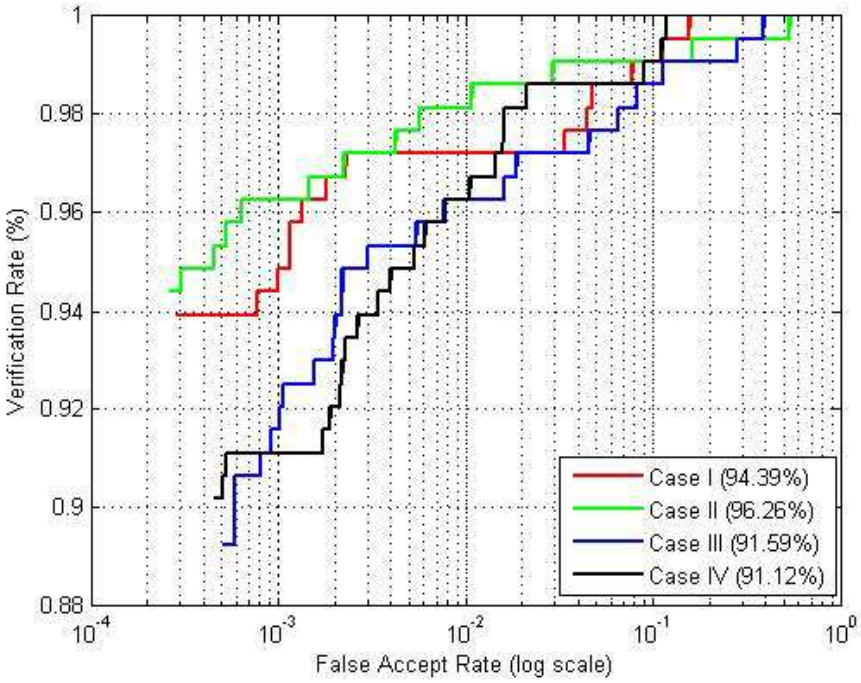


Figure G.7: ROC curves in four cases of our method based on N-GFs. The legend shows TAR at 0.1% FAR.

Algorithm	True Accept Rate			
	I	II	III	IV
Alg.1 (E_{pkn}) [Faltemier *et al.* 2008b]	79.0%	81.3%	54.2%	53.3%
Alg.1 (E_{minmax}) [Faltemier *et al.* 2008b]	**99.5%**	97.7%		
Alg.2 (SI) [Huang *et al.* 2011f]	91.1%	89.7%	83.2%	81.8%
Alg.2 (eLBP) [Huang *et al.* 2011c]	94.4%	95.3%	79.0%	78.0%
Alg.2 (Range PFI) [Huang *et al.* 2011d]	93.5%	94.4%	68.7%	69.2%
Alg.2 (Text. PFI) [Huang *et al.* 2011d]	96.7%	96.7%	93.0%	93.5%
Alg.3 [Gokberk *et al.* 2006]	38.1%	41.0%	31.4%	34.1%
Alg.4 [Kakadiaris *et al.* 2007b]	98.1%	**98.1%**	**95.8%**	**95.8%**
Alg.5 (N-LBPs)	96.7%	**98.1%**	92.1%	91.6%
Alg.5 (N-GFs)	94.4%	96.3%	91.6%	91.1%
Alg.5 (N-LGBPs)	**97.2%**	97.7%	**94.4%**	**94.9%**

Table G.4: Comparison of TAR at 0.1% FAR of our method (Alg. 5) and the state-of-the-art methods.

Linear Discriminant Analysis (LDA), AdaBoost, manifold learning, etc. will be discussed for N-GFs and N-LGBPs descriptors.

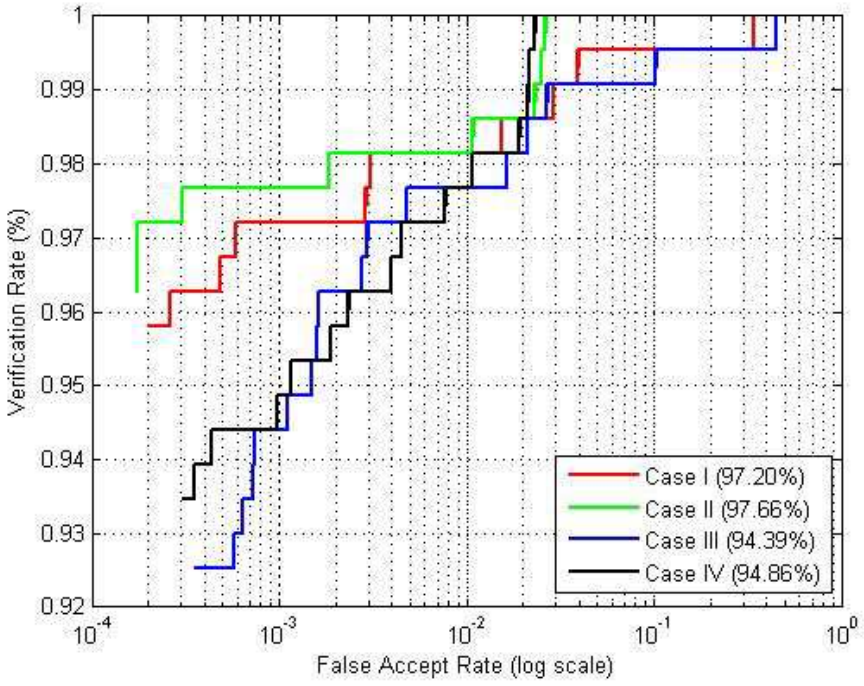


Figure G.8: ROC curves in four cases of our method based on N-LGBPs. The legend shows TAR at 0.1% FAR.

Algorithm	Equal Error Rate			
	I	II	III	IV
Alg.1 (E_{pkn}) [Faltemier *et al.* 2008b]	1.2%	1.0%	1.4%	1.1%
Alg.1 (E_{minmax}) [Faltemier *et al.* 2008b]	**0.2%**	**0.5%**	1.3%	0.9%
Alg.2 (SI) [Huang *et al.* 2011f]	2.7%	3.7%	4.2%	4.5%
Alg.2 (eLBP) [Huang *et al.* 2011c]	3.7%	3.3%	4.2%	4.2%
Alg.2 (Range PFI) [Huang *et al.* 2011d]	4.1%	2.8%	4.7%	4.6%
Alg.2 (Text. PFI) [Huang *et al.* 2011d]	2.7%	2.8%	3.3%	2.8%
Alg.3 [Gokberk *et al.* 2006]	11.6%	11.8%	12.0%	12.2%
Alg.4 [Kakadiaris *et al.* 2007b]	0.8%	0.8%	**0.8%**	**0.8%**
Alg.5 (N-LBPs)	**1.4%**	**1.4%**	**1.4%**	**1.9%**
Alg.5 (N-GFs)	2.8%	**1.4%**	2.8%	**1.9%**
Alg.5 (N-LGBPs)	**1.4%**	**1.4%**	1.9%	**1.9%**

Table G.5: Comparison of EER of our method (Alg. 5) and the state-of-the-art methods.

APPENDIX H

Publications

During my Ph.D studying, I have published nine publications in international conferences, and three journal papers are currently under review.

Published International Conference Papers :

1. RC. Veltkamp, S. van Jole, H. Drira, B Ben Amor, M. Daoudi, **H. Li**, L. Chen, P. Claes, D. Smeets, J. Hermans: SHREC'11 Track: 3D Face Models Retrieval, Eurographics 2011 Workshop on 3D Object Retrieval (3DOR), pp. 89-95, Llandudno, UK, 2011;

2. **H. Li**, D. Huang, J. M. Morvan, L. Chen: Learning Weighted Sparse Representation of Encoded Facial Normal Information for Expression-robust 3D Face Recognition, IEEE International Joint Conference on Biometrics (IJCB), pp. 1-7, Washington, DC, USA, 2011;

3. **H. Li**, P. Lemaire, J.-M. Morvan, L. Chen: Expression-robust 3D Face Recognition via Mesh-based Histograms of Multiple-order Surface Differential Quantities, IEEE International Conference on Image Processing (ICIP), pp. 3053-3056, Brussels, Belgium, 2011.

4. **H. Li**, J.-M. Morvan, L. Chen: 3D Facial Expression Recognition based on Histograms of Surface Differential Quantities, Advances Concepts for Intelligent Vision Systems (ACIVS), pp. 483-494, Ghent, Belgium, 2011.

5. **H. Li**, D. Huang, L. Chen, Y. Wang, J. M. Morvan: A Group of Facial Normal Descriptors for Recognizing 3D Identical Twins, Fifth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 271-277, Washington, DC, USA, 2012;

6. **H. Li**, L. Chen, D. Huang, Y. Wang, J. M. Morvan: 3D Facial Expression Recognition via Multiple Kernel Learning of Multi-scale Local Normal Patterns, 21st IEEE International Conference on Pattern Recognition (ICPR), pp. 2577-2580, Tsukuba Science City, Japan, 2012;

7. W. Zeng, **H. Li**, L. Chen, J. M. Morvan, X. David Gu: An Automatic 3D Expression Recognition Framework based on Sparse Representation of Conformal Images, 10st IEEE International Conference on Automatic Face and Gesture Recognition (FG), Shanghai, China, 2013;

8. X. Wang, **H. Li**, C. E. Bichot, S. Masnou, L. Chen: A Graph-cut Approach to Image Segmentation using Affinity Graph based on l_0 Sparse Representation of Features, IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 2013 (accept);

9. X. Wang, **H. Li**, C. E. Bichot, S. Masnou, L. Chen: Sparse Coding and Mid-level Super-pixel Feature for l_0-Graph based Unsupervised Image Segmentation, 15st International Conference on Computer Analysis of Images and Patterns (CAIP), York, UK, 2013 (accept).

Submitted International Journal Manuscripts:

1. **H. Li**, W. Zeng, J. M. Morvan, L. Chen, X. David Gu: Surface Meshing with Curvature Convergence, Submitted to IEEE Transactions on Visualization and Computer Graphics (TVCG), Submission Date: Nov. 26th, 2012, Submission Number: TVCG-2012-11-0263, Manuscript Status: Under Review.

2. **H. Li**, D. Huang, J. M. Morvan, L. Chen, Y. Wang: Expression-Robust 3D Face Recognition via Weighted Sparse Representation of Multi-Scale and Multi-Component Local Normal Patterns, Submitted to Neurocomputing, Submission Date: March 24th, 2013, Submission Number: NEUCOM-S-13-00419, Manuscript Status: Under Review.

3. **H. Li**, D. Huang, J. M. Morvan, Y. Wang, L. Chen: Towards 3D Face Recognition in the Real: A Registration Free Local Feature Matching Approach using Mesh-based Histograms of Multiple Order Surface Differential Quantities, Submitted to International Journal of Computer Vision (IJCV), Submission Date: April 26th, 2013, Submission Number: VISI-S-13-00373, Manuscript Status: Under Review.

Bibliography

[3DM 2011] 3dmd: 4d capture system url. http://www.di3d.com/products/4dsystems, 2011. 14

[A. Maalej 2010] M. Daoudi A. Srivastava S. Berretti A. Maalej B. Ben Amor. *Local 3D Shape Analysis for Facial Expression Recognition*. In ICPR, pages 4129–4132, 2010. 207, 208

[Abate *et al.* 2005] A. F. Abate, M. Nappi, S. Ricciardi and G. Sabatino. *Fast 3D face recognition based on normal map*. In Proc. IEEE Int. Conf. on Image Processing, 2005. 69

[Abate *et al.* 2006] A. F. Abate, M. Nappi, S. Ricciardi and G. Sabatino. *Multi-Modal Face Recognition by Means of Augmented Normal Map and PCA*. In Proc. IEEE Int. Conf. on Image Processing, 2006. 69

[Abate *et al.* 2007a] A. F. Abate, M. Nappi, S. Ricciardi and G. Sabatino. *Fast 3D Face Alignment and Improved Recognition Through Pyramidal Normal map Metric*. In Proc. IEEE Int. Conf. on Image Processing, 2007. 69

[Abate *et al.* 2007b] Andrea F. Abate, Michele Nappi, Daniel Riccio and Gabriele Sabatino. *2D and 3D face recognition: A survey*. Pattern Recognition Letters, vol. 28, no. 14, pages 1885–1906, 2007. 98

[Abate *et al.* 2009] A. F. Abate, M. D. Marsico, S. Ricciardi and D. Riccio. *Normal maps vs. visible images: Comparing classifiers and combining modalities*. Journal of Visual Languages and Computing, vol. 20, no. 3, pages 156–168, 2009. 69, 92

[Ahonen *et al.* 2004a] T. Ahonen, A. Hadid and M. Pietikäinen. *Face Recognition with Local Binary Patterns*. In Proc. Int. Conf. European Conference on Computer Vision, 2004. xvi, 70, 71

[Ahonen *et al.* 2004b] Timo Ahonen, Abdenour Hadid and Matti Pietikainen. *Face Recognition with Local Binary Patterns*. In European Conference on Computer Vision (ECCV), pages 469–481, 2004. 242

[Ahonen *et al.* 2006] 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, pages 2037–2041, Dec. 2006. 9

[Al-Osaimi *et al.* 2008] F. R. Al-Osaimi, M. Bennamoun and A. S. Mian. *Integration of local and global geometrical cues for 3D face recognition*. Pattern Recognition, vol. 41, no. 3, pages 1030–1040, 2008. 92

[Al-Osaimi *et al.* 2009] F. Al-Osaimi, M. Bennamoun and A. Mian. *An Expression Deformation Approach to Non-rigid 3D Face Recognition*. Int. J. Comput. Vision, vol. 81, no. 3, pages 302–316, March 2009. xv, 30, 38, 39, 92, 94, 137

[Alliez *et al.* 2003] Pierre Alliez, Éric Colin de Verdière, Olivier Devillers and Martin Isenburg. *Isotropic Surface Remeshing*. In Proceedings of the Shape Modeling International, pages 49–59, 2003. 25, 144, 147

[Alliez *et al.* 2008] P Alliez, G Ucelli, C Gotsman and M Attene. *Recent Advances in Remeshing of Surfaces*. In Shape Analysis and Structuring, Mathematics and Visualization, pages 53–82, 2008. 25, 144, 147

[Alyuz *et al.* 2008] N. Alyuz, B. Gokberk and L. Akarun. *A 3D Face Recognition System for Expression and Occlusion Invariance*. In 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, pages 1–7, 2008. 127, 132

[Alyuz *et al.* 2010] 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, pages 425–440, 2010. xvi, 49, 51, 53, 56, 61, 64, 68, 69, 92, 94, 128, 132, 137

[Alyüz *et al.* 2012a] Nese Alyüz, Berk Gökberk and Lale Akarun. *Adaptive Registration for Occlusion Robust 3D Face Recognition*. In ECCV Workshops (3), pages 557–566, 2012. 64

[Alyüz *et al.* 2012b] Nese Alyüz, Berk Gökberk, Luuk J. Spreeuwers, Raymond N. J. Veldhuis and Lale Akarun. *Robust 3D face recognition in the presence of realistic occlusions*. In ICB, pages 111–118, 2012. 64

[Alyüz *et al.* 2013] Nese Alyüz, Berk Gökberk and Lale Akarun. *3-D Face Recognition Under Occlusion Using Masked Projection*. IEEE Transactions on Information Forensics and Security, vol. 8, no. 5, pages 789–802, 2013. 64

[Amberg *et al.* 2008] B. Amberg, R. Knothe and T. Vetter. *Expression invariant 3D face recognition with a Morphable Model*. In IEEE International Conference on Automatic Face Gesture Recognition, pages 1–6, Sep. 2008. xv, 30, 40, 41

[Amenta & Bern 1998] Nina Amenta and Marshall Bern. *Surface reconstruction by Voronoi filtering*. In Proceedings of the fourteenth annual symposium on Computational geometry (SCG '98), pages 39–48, 1998. 24, 25, 144, 145

[Amenta *et al.* 2000] N. Amenta, S. Choi, T. K. Dey and N. Leekha. *A simple algorithm for homeomorphic surface reconstruction*. In Proceedings of the sixteenth annual symposium on Computational geometry (SCG '00), pages 213–222, 2000. 24, 144

[Amor *et al.* 2006a] B.B. Amor, M. Ardabilian and Liming Chen. *Enhancing 3D Face Recognition By Mimics Segmentation*. In 6th International Conference on Intelligent Systems Design and Applications, pages 150–155, 2006. 105

[Amor *et al.* 2006b] Boulbaba Ben Amor, Mohsen Ardabilian and Liming Chen. *Enhancing 3D Face Recognition By Mimics Segmentation*. In 6th International Conference on Intelligent Systems Design and Applications (ISDA), pages 150–155, 2006. 68, 105

[Amor *et al.* 2006c] Boulbaba Ben Amor, Mohsen Ardabilian and Liming Chen. *New Experiments on ICP-Based 3D Face Recognition and Authentication*. In International Conference on Pattern Recognition, 2006. 104, 105

[Anil K. Jain 2007] Arun A. Ross Anil K. Jain Patrick Flynn. Handbook of biometrics. Springer, 2007. 4

[Anil K. Jain 2011] Karthik Nandakumar Anil K. Jain Arun A. Ross. Introduction to biometrics. Springer, 2011. 4

[B. Gong & Tang 2009] J. Liu B. Gong Y. Wang and X. Tang. *Automatic facial expression recognition on a single 3d face by exploring shape deformation*. In In Int. Conf. on Multimedia, pages 569–572, 2009. 207, 208, 213, 216

[Bai 1992] Zhengguo Bai. An introduction to riemannian geometry (in chinese). High Education Press, 1992. 175

[Ballihi *et al.* 2012] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava and D. Aboutajdine. *Boosting 3-D-Geometric Features for Efficient Face Recognition and Gender Classification*. Information Forensics and Security, IEEE Transactions on, vol. 7, no. 6, pages 1766–1779, 2012. 45, 46, 137

[Beeler *et al.* 2010] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner and Markus Gross. *High-quality single-shot capture of facial geometry*. ACM Transactions on Graphics (TOG), vol. 29, no. 4, page 40, 2010. xiii, 14, 15

[Belhumeur *et al.* 1997] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 711–720, Jul. 1997. 8

[Belongie *et al.* 2002] Serge Belongie, Jitendra Malik and Jan Puzicha. *Shape matching and object recognition using shape contexts*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 4, pages 509–522, 2002. 63

[Ben Amor *et al.* 2008] B. Ben Amor, M. Ardabilian and Liming Chen. *Toward a region-based 3D face recognition approach*. In IEEE International Conference on Multimedia and Expo, pages 101–104, 2008. 68, 105, 106

[Benedikt *et al.* 2010] Lanthao Benedikt, Darren Cosker, Paul L Rosin and David Marshall. *Assessing the uniqueness and permanence of facial actions for use in biometric applications*. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, no. 3, pages 449–460, 2010. 14

[Berretti *et al.* 2006] Stefano Berretti, Alberto Del Bimbo and Pietro Pala. *Description and retrieval of 3D face models using iso-geodesic stripes*. In Proceedings of the 8th ACM international workshop on Multimedia information retrieval, pages 13–22, 2006. 48

[Berretti *et al.* 2008] S. Berretti, A. Del Bimbo and P. Pala. *Analysis and retrieval of 3D facial models using iso-geodesic stripes*. In Content-Based Multimedia Indexing, 2008. CBMI 2008. International Workshop on, pages 257–264, 2008. 48

[Berretti *et al.* 2010a] 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, pages 2162–2177, 2010. xv, 48, 49, 68, 105

[Berretti *et al.* 2010b] S. Berretti, A.D. Bimbo, P. Pala, B.B. Amor and M. Daoudi. *A Set of Selected SIFT Features for 3D Facial Expression Recognition*. In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 4125–4128, 2010. xi, 217, 218, 221, 222, 223, 226, 233, 237, 238, 239

[Berretti *et al.* 2011a] Stefano Berretti, Boulbaba Ben Amor, Mohamed Daoudi and Alberto del Bimbo. *3D facial expression recognition using SIFT descriptors of automatically detected keypoints*. Vis. Comput., vol. 27, no. 11, pages 1021–1036, 2011. 237, 239

[Berretti *et al.* 2011b] Stefano Berretti, Alberto del Bimbo and Pietro Pala. *3D partial face matching using local shape descriptors*. In Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, pages 65–71, 2011. 50, 63, 65, 173

[Berretti *et al.* 2013] Stefano Berretti, Alberto Del Bimbo and Pietro Pala. *Sparse Matching of Salient Facial Curves for Recognition of 3D Faces with Missing Parts*. Information Forensics and Security, IEEE Transactions on, vol. 8, no. 2, pages 374–389, 2013. 65, 173

[Besl & McKay 1992] P.J. Besl and Neil D. McKay. *A method for registration of 3-D shapes*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 14, no. 2, pages 239–256, 1992. 104

[Beumier & Acheroy 1999] Charles Beumier and Marc Acheroy. *3D Facial Surface Acquisition by Structured Light*. In In International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, pages 103–106, 1999. 14

[Bicego *et al.* 2006] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli. *On the use of SIFT features for face authentication*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, volume 0, pages 35–40, 2006. 9, 105

[Blanz & Vetter 1999] V. Blanz and T. Vetter. *A morphable model for the synthesis of 3D faces*. In Annual Conference on Computer Graphics, pages 187–194, 1999. 40

[Blanz & Vetter 2003] Volker Blanz and Thomas Vetter. *Face Recognition Based on Fitting a 3D Morphable Model*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 9, pages 1063–1074, 2003. 9

[Boissonnat & Oudot 2005] Jean-Daniel Boissonnat and Steve Oudot. *Provably good sampling and meshing of surfaces*. Graph. Models, vol. 67, no. 5, pages 405–451, 2005. 24, 25, 144, 145

[Borrelli *et al.* 2003] V. Borrelli, F. Cazals and J. M. Morvan. *On the angular defect of triangulations and the pointwise approximation of curvatures*. Computer Aided Geometric Design, vol. 20, pages 319–341, 2003. 200

[Bowyer *et al.* 2005] K. W. Bowyer, K. Chang and P. Flynn. *Adaptive rigid multi-region selection for handling expression variation in 3D face recognition*. In Proc. Int. Conf. Computer Vision and Pattern Recognition, 2005. 49, 50, 54, 92

[Bowyer *et al.* 2006] 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, pages 1–15, 2006. 4, 13, 14, 16, 17, 67, 98, 103

[Bronstein *et al.* 2003] Alexander M. Bronstein, Michael M. Bronstein and Ron Kimmel. *Expression-invariant 3D face recognition*. In 4th international conference on Audio- and video-based biometric person authentication, pages 62–70, 2003. 43, 64

[Bronstein *et al.* 2004a] A. M. Bronstein, M. M. Bronstein, E. Gordon and R. Kimmel. *Fusion of 2D and 3D data in three-dimensional face recognition*. In IEEE International Conference on Image Processing, volume 1, pages 87–90, Oct. 2004. 44

[Bronstein *et al.* 2004b] AlexanderM. Bronstein, MichaelM. Bronstein, Alon Spira and Ron Kimmel. *Face Recognition from Facial Surface Metric*. In European Conference on Computer Vision, pages 225–237, 2004. 44

[Bronstein *et al.* 2005] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Three-dimensional face recognition*. International Journal of Computer Vision, vol. 64, no. 1, pages 5–30, 2005. xv, 44

[Bronstein *et al.* 2006] Alexander M. Bronstein, Michael M. Bronstein and Ron Kimmel. *Robust expression-invariant face recognition from partially missing data*. In IN PROC. ECCV, LECTURE NOTES ON COMPUTER SCIENCE, pages 396–408. Springer, 2006. 44, 64

[Bronstein *et al.* 2007] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Expression-invariant representations of faces*. IEEE Transactions on Image Processing, vol. 16, no. 1, pages 188–197, Jan. 2007. 43, 44, 45, 68, 104, 105

[Carmo 1976] Manfredo P. Do Carmo. Differential geometry of curves and surfaces. Prentice-Hall, inc., Englewood Cliffs, New Jersey, 1976. 175

[Chang & Lin 2001] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM : a library for support vector machines*. In `http://www.csie.ntu.edu.tw/~cjlin/libsvm`, 2001. 214

[Chang & Lin 2011] C. C. Chang and C. J. Lin. *LIBSVM: A library for support vector machines*. ACM Trans. IST, 2011. 221

[Chang *et al.* 2006] 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, pages 1695–1700, 2006. xv, 49, 50, 51, 54, 68, 70, 105, 137

[Cheng *et al.* 2004] S. W. Cheng, T. K. Dey, E. A. Ramos and T. Ray. *Sampling and meshing a surface with guaranteed topology and geometry*. In Proc. 20th Annu. Sympos. Comput. Geom., pages 280–289, 2004. 144, 147

[Cheng *et al.* 2012] Siu-Wing Cheng, Tamal K. Dey and Jonathan R. Shewchuk. Delaunay mesh generation. CRC Press, 2012. 147

[Chu *et al.* 2013] Dat Chu, Shishir K. Shah and Ioannis A. Kakadiaris. *3D Face Recognition for Partial Data using Semi-Coupled Dictionary Learning*. In IEEE 10th International Conference on Automatic Face and Gesture Recognition, FG '13., 2013. 37, 38, 64, 173

[Chua & Jarvis 1997] ChinSeng Chua and Ray Jarvis. *Point Signatures: A New Representation for 3D Object Recognition*. International Journal of Computer Vision, vol. 25, no. 1, pages 63–85, 1997. 50, 62

[Chua *et al.* 2000] C. Chua, F. Han and Y. Ho. *3D human face recognition using point signature*. In IEEE International Conference on Automatic Face and Gesture Recognition, pages 233–238, 2000. 62, 104

[Cohen-Steiner & Morvan 2003a] David Cohen-Steiner and Jean Marie Morvan. *Restricted delaunay triangulations and normal cycle*. In Proceedings of the nineteenth Annual Symposium on Computational Geometry (SCG '03), pages 312–321, 2003. 25, 26, 144, 145, 146, 153, 161

[Cohen-Steiner & Morvan 2003b] David Cohen-Steiner and J.M. Morvan. *Restricted delaunay triangulations and normal cycle*. In Proceedings of the nineteenth annual symposium on Computational geometry, pages 312–321, 2003. 208, 209

[Cohen-Steiner & Morvan 2006] David Cohen-Steiner and Jean-Marie Morvan. *Second fundamental measure of geometric sets and local approximation of curvatures*. J. Differential Geom., vol. 74, no. 3, pages 363–394, 2006. 25, 144, 145

[Colombo *et al.* 2011a] Alessandro Colombo, Claudio Cusano and Raimondo Schettini. *Three-dimensional occlusion detection and restoration of partially occluded faces*. Journal of Mathematical Imaging and Vision, vol. 40, no. 1, pages 105–119, 2011. 20, 100, 129, 132

[Colombo *et al.* 2011b] Alessandro Colombo, Claudio Cusano and Raimondo Schettini. *UMB-DB: A Database of Partially Occluded 3D Faces*. In IEEE International Conference on Computer Vision Workshops, pages 2113–2119, 2011. 17

[Conde *et al.* 2006] Cristina Conde, LicesioJ. Rodríguez-Aragón and Enrique Cabello. *Automatic 3D Face Feature Points Extraction with Spin Images*. In Image Analysis and Recognition, volume 4142, pages 317–328, 2006. 50, 62

[Cook *et al.* 2006] Jamie Cook, Vinod Chandran and Clinton Fookes. *3D Face Recognition using Log-Gabor Templates*. In British Machine Vision Conference (BMVC), pages 769–778, 2006. 50, 57, 92, 243

[Cook *et al.* 2007] Jamie Cook, Vinod Chandran and Sridha Sridharan. *Multiscale Representation for 3-D Face Recognition*. IEEE Transactions on Information Forensics and Security, vol. 2, no. 3-2, pages 529–536, 2007. 50, 57

[Cosker *et al.* 2011] Darren Cosker, Eva Krumhuber and Adrian Hilton. *A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling*. In 2011 IEEE International Conference on Computer Vision (ICCV), pages 2296–2303. IEEE, 2011. 14

[Cotter 2010] Shane F. Cotter. *Sparse Representation for accurate classification of corrupted and occluded facial expressions*. In ICASSP, pages 838–841, 2010. 226

[Dai *et al.* 2007] Junfei Dai, Wei Luo, Miao Jin, Wei Zeng, Ying He, Shing-Tung Yau and Xianfeng Gu. *Geometric accuracy analysis for discrete surface approximation*. Computer Aided Geometric Design, vol. 24, no. 6, pages 323–338, 2007. 24, 144, 145

[Dantcheva *et al.* 2012] A. Dantcheva, Cunjian Chen and A. Ross. *Can facial cosmetics affect the matching accuracy of face recognition systems?* In 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pages 391–398, 2012. xiii, 11

[Darom & Keller 2012] Tal Darom and Yosi Keller. *Scale-Invariant Features for 3-D Mesh Models*. Image Processing, IEEE Transactions on, vol. 21, no. 5, pages 2758–2769, 2012. 65

[Desbrun *et al.* 2002] Mathieu Desbrun, Mark Meyer and Pierre Alliez. *Intrinsic Parameterizations of Surface Meshes*. Computer Graphics Forum, vol. 21, no. 3, pages 209–218, 2002. 148

[Dey & Levine 2007] T. K. Dey and J. A. Levine. *Delaunay meshing of isosurfaces.* In Proc. Shape Modeling International, pages 241–250, 2007. 144, 147

[Dey *et al.* 2005] T. K. Dey, G. Li and T. Ray. *Polygonal surface remeshing with Delaunay refinement*. In Proc. 14th Intl. Meshing Roundtable., pages 343–361, 2005. 144, 147

[D.G.Lowe 2004] D.G.Lowe. *Distinctive image features from scale invariant keypoints*. IJCV, pages 91–110, 2004. 208

[DI4 2011] Di4d: 4d capture system url. http://www.di3d.com/products/4dsystems, 2011. 14

[Dibeklíouglu *et al.* 2009] H. Dibeklíouglu, B. Gokberk and Lale Akarun. *Nasal Region-Based 3D Face Recognition under Pose and Expression Variations*. In Advances in Biometrics, volume 5558, pages 309–318, 2009. 100, 127, 128, 132

[Donoho 2006] DL Donoho. *For most large underdetermined systems of equations, the minimal l_1-norm near-solution approximates the sparsest near-solution*. Communications on Pure and Applied Mathematics, vol. 59, no. 7, 2006. 232

[Drira *et al.*] Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi and Anuj Srivastava. *Pose and Expression-Invariant 3D Face Recognition using Elastic Radial Curves*. In British Machine Vision Conference. 45, 46

[Drira *et al.* 2013] Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi and Rim Slama. *3D Face Recognition Under Expressions, Occlusions and Pose Variations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, no. PrePrints, 2013. xv, 29, 45, 46, 47, 48, 64, 65, 68, 100, 103, 104, 129, 132, 173

[Du *et al.* 1999] Qiang Du, Vance Faber and Gunzburger Max. *Centroidal Voronoi Tessellations: Applications and Algorithms*. Society for Industrial and Applied Mathematics Review, vol. 41, pages 637–676, 1999. 147

[Du *et al.* 2002] Qiang Du, Gunzburger Max and Lili Ju. *Constrained Centroidal Voronoi Tessellations for Surfaces*. Society for Industrial and Applied Mathematics J. Sci. Comput., vol. 24, pages 1488–1506, 2002. 147

[Edwards *et al.* 1998] G. J. Edwards, T. F. Cootes and C. J. Taylor. *Face recognition using active appearance models*. In European Conference on Computer Vision, pages 581–595, 1998. 9

[Ekman 1972] P. Ekman. *Universals and cultural differences in facial expressions of emotion*. In In Nebraska Symposium on Motivation, pages 207–283, 1972. 207

[et al 2010] M.C.Fabry et al. *Feature detection on 3d face surfaces for pose normalisation and recognition*. In BTAS, 2010. 208

[Faltemier *et al.* 2006] . Faltemier, K. Bowyer and P. Flynn. *3D face recognition with region committee voting*. In Proc. 3rd Int. Symp. 3D Data Processing, Visualization Transmission, page 318¨C325, 2006. 51

[Faltemier *et al.* 2008a] 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, pages 62–73, 2008. xv, 49, 50, 52, 54, 56, 68, 70, 92, 94, 99, 103, 105, 106, 121, 137

[Faltemier *et al.* 2008b] Timothy C. Faltemier, Kevin W. Bowyer and Patrick J. Flynn. *A Region Ensemble for 3D Face Recognition*. IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pages 62–73, 2008. 251, 253, 254

[Faltemier 2007] T. C. Faltemier. *Flexible and robust 3D face recognition*. PhD thesis, The University of Notre Dame, 2007. 13

[Fang *et al.* 2011] T. Fang, X. Zhao, O. Ocegueda, S.K. Shah and I.A. Kakadiaris. *3D facial expression recognition: A perspective on promises and challenges*. In IEEE

International Conference on Automatic Face Gesture Recognition and Workshops(FG), pages 603 –610, 2011. 217, 226

[Fang *et al.* 2012] Tianhong Fang, Xi Zhao, Omar Ocegueda, Shishir K. Shah and Ioannis A. Kakadiaris. *3D/4D facial expression analysis: An advanced annotated face model approach*. Image and Vision Computing, 2012. in press. 226

[Fasel & Luettin 1999] Beat Fasel and Juergen Luettin. *Automatic Facial Expression Analysis: A Survey*. Pattern Recognition, vol. 36, no. 1, pages 259–275, 1999. 225

[Fasel & Luettin.J. 2003] B. Fasel and Luettin.J. *Automatic facial expression analysis: a survey.* In Pattern Recognition, pages 259–275, 2003. 207

[Federer 1983] H. Federer. Geometric measure theory. Springer, 1983. 144

[Fischler & Bolles 1981] Martin A. Fischler and Robert C. Bolles. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Commun. ACM, vol. 24, no. 6, pages 381–395, June 1981. 56

[Floater & Hormann 2005] Michael S. Floater and Kai Hormann. *Surface parameterization: a tutorial and survey*. In Advances in Multiresolution for Geometric Modelling, pages 157–186, 2005. 148

[Frankot *et al.* 1988] Robert T. Frankot, Rama Chellappa and Senior Member. *A Method for enforcing integrability in shape from shading algorithms*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, pages 439–451, 1988. 15

[Frome *et al.* 2004] Andrea Frome, Daniel Huber, Ravi Kolluri, Thomas Bülow and Jitendra Malik. *Recognizing objects in range data using regional point descriptors*. In Computer Vision-ECCV 2004, pages 224–237. Springer, 2004. 62

[Fu 1989] J. Fu. *Monge-Amp¨¨re functions 1.* Indiana University Mathematics, vol. 38, pages 745–771, 1989. 144

[Funke & Ramos 2002] Stefan Funke and Edgar A. Ramos. *Smooth-Surface Reconstruction in Near Linear Time*. In 13th ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 781–790, 2002. 153

[GATZKE & Grimm 2006] T. D. GATZKE and Cindy M. Grimm. *Estimating Curvature on Triangular Meshes*. International Journal of Shape Modeling, vol. 12, pages 1–28, 2006. 107

[G.G.Gordon 1992] G.G.Gordon. *Face Recognition Based on Depth and Curvature Features*. In Proc. Int. Conf. Computer Vision and Pattern Recognition, 1992. 69

[Gokberk *et al.* 2006] B. Gokberk, M. O. Irfanoglu and L. Akarun. *3D shape-based face representation and feature extraction for face recognition*. Image and Vision Computing, vol. 24, no. 8, pages 857–869, 2006. 52, 69, 92, 251, 253, 254

[Gokberk *et al.* 2008] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun and B. Sankur. *Representation Plurality and Fusion for 3-D Face Recognition*. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 38, no. 1, pages 155–173, 2008. 52, 68, 103

[Goldfeather & Interrante 2004a] J. Goldfeather and V. Interrante. *A novel cubic-order algorithm for approximating principal direction vectors*. ACM Trans. Graph., pages 45–63, 2004. 209

[Goldfeather & Interrante 2004b] Jack Goldfeather and Victoria Interrante. *A novel cubic-order algorithm for approximating principal direction vectors*. ACM Trans. Graph., vol. 23, no. 1, pages 45–63, 2004. 107

[Gong *et al.* 2009] Boqing Gong, Yueming Wang, Jianzhuang Liu and Xiaoou Tang. *Automatic facial expression recognition on a single 3D face by exploring shape deformation*. In Proceedings of the 17th ACM international conference on Multimedia (MM), pages 569–572, 2009. xi, 217, 218, 222, 223, 226, 237, 238, 239

[Gordon 1992] G. G. Gordon. *Face recognition based on depth and curvature features*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 808–810, Jun. 1992. xvi, 13, 50, 60, 61

[Grother *et al.*] P. J. Grother, G. W. Quinn and P. J. Phillips. *Report on the Evaluation of 2D Still-Image Face Recognition Algorithms*. NIST Interagency/Internal Report (NISTIR) - 7709, 2010. 241

[Gu & Vemuri 2004] Xianfeng Gu and Baba C. Vemuri. *Matching 3D Shapes Using 2D Conformal Representations*. In MICCAI (1), pages 771–780, 2004. 226, 229

[Gu & Yau 2003] Xianfeng Gu and Shing-Tung Yau. *Global conformal surface parameterization*. In Proceedings of the Symposium on Geometry Processing (SGP '03), pages 127–137, 2003. 148

[Gu *et al.* 2002] Xianfeng Gu, Steven J. Gortler and Hugues Hoppe. *Geometry images*. In Proceedings of the 29th annual conference on Computer graphics and interactive techniques (SIGGRAPH '02), pages 355–361, 2002. 33

[Gu *et al.* 2004] Xianfeng Gu, Yalin Wang, Tony F. Chan, Paul M. Thompson and Shing-Tung Yau. *Genus zero surface conformal mapping and its application to brain surface mapping*. IEEE Transaction on Medical Imaging, vol. 23, no. 8, pages 949–958, 2004. 148

[Guo *et al.* 2012] Huimin Guo, Ruiping Wang, Jonghyun Choi and L.S. Davis. *Face verification using sparse representations*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), pages 37–44, 2012. 95

[Hajati *et al.* 2012] Farshid Hajati, Abolghasem A. Raie and Yongsheng Gao. *2.5D face recognition using Patch Geodesic Moments*. Pattern Recognition, vol. 45, no. 3, pages 969–982, 2012. 128, 132

[Haker *et al.* 2000] S. Haker, S. Angenent, A. Tannenbaum, R. Kikinis, G. Sapiro and M. Halle. *Conformal surface parameterization for texture mapping*. IEEE Transactions on Visualization and Computer Graphics, vol. 6, no. 2, pages 181–189, 2000. 148

[Hamilton 1982] Richard Hamilton. *Three manifolds with positive Ricci curvature*. Journal of Differential Geometry, vol. 17, pages 255–306, 1982. 229

[Hansen *et al.* 2010] Mark F. Hansen, Gary A. Atkinson, Lyndon N. Smith and Melvyn L. Smith. *3D face reconstructions from photometric stereo using near infrared and visible light*. Computer Vision and Image Understanding, vol. 114, no. 8, pages 942–951, 2010. xiv, 15

[He *et al.* 2005] X. He, S. Yan, Y. Hu, P. Niyogi and H. Zhang. *Face recognition using Laplacianfaces*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pages 328–340, Mar. 2005. 9

[Heseltine *et al.* 2004] T. Heseltine, N. Pears and J. Austin. *Three-dimensional face recognition: an eigensurface approach*. In International Conference on Image Processing, pages 1421–1424, 2004. 105

[Hietmeyer 2000] R. Hietmeyer. *Biometric Identification promises fast and secure processing of airline passengers*. Internation Civil Aviation Organization Journal, vol. 55, no. 9, pages 10–11, 2000. xiii, 5, 6

[Hildebrandt *et al.* 2005] Klaus Hildebrandt, Konrad Polthier and Max Wardetzky. *On the convergence of metric and geometric properties of polyhedral surfaces*. Geometriae Dedicata, vol. 123, pages 89–112, 2005. 170

[Hoffman & Jain 1987] Richard Hoffman and Anil K. Jain. *Segmentation and Classification of Range Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, no. 5, pages 608–620, 1987. 74, 244

[Hormann *et al.* 1999] K. Hormann, G. Greiner and S. Campagna. *Hierarchical Parametrization of Triangulated Surfaces*. In Proceedings of Vision, Modeling, and Visualization, pages 219–226, 1999. 148

[Huang *et al.* 2003] Peisen S Huang, Chengping Zhang and Fu-Pen Chiang. *High-speed 3-D shape measurement based on digital fringe projection*. Optical Engineering, vol. 42, no. 1, pages 163–168, 2003. 14

[Huang *et al.* 2010] 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 Proc. Int. Conf. Biometrics: Theory Applications and Systems, 2010. 68, 69, 90, 92

[Huang *et al.* 2011a] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *A Novel Geometric Facial Representation based on Multi-Scale Extended Local Binary Patterns*. In Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2011. 50, 59, 64, 90, 92, 113, 137

[Huang *et al.* 2011b] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *A novel geometric facial representation based on multi-scale extended local binary patterns*. In Proc. Int. Conf. Automatic Face and Gesture Recognition, 2011. 90

[Huang *et al.* 2011c] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *A Novel Geometric Facial Representation based on Multi-Scale Extended Local Binary Patterns*. In Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2011. 243, 251, 253, 254

[Huang *et al.* 2011d] D. Huang, W. Ben Soltana, M. Ardabilian, Y. Wang and L. Chen. *Textured 3D face recognition using biological vision-based facial representation and optimized weighted sum fusion*. In Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, 2011. 92, 251, 253, 254

[Huang *et al.* 2011e] 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 Proc. Int. Conf. on Biometrics: Theory, Applications and Systems, 2011. 59, 103, 105, 137

[Huang *et al.* 2011f] 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 Proc. Int. Conf. on Biometrics: Theory, Applications and Systems, 2011. 251, 253, 254

[Huang *et al.* 2011g] Di Huang, Karima Ouji, Mohsen Ardabilian, Yunhong Wang and Liming Chen. *3D Face Recognition Based on Local Shape Patterns and Sparse Representation Classifier*. In Advances in Multimedia Modeling:17th International Multimedia Modeling Conference (MMM), pages 206–216, 2011. 70

[Huang *et al.* 2011h] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang and Liming Chen. *Local Binary Patterns and Its Application to Facial Image Analysis: A Survey*. IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 41, no. 6, pages 765–781, 2011. 70

[Huang *et al.* 2012] Di Huang, Mohsen Ardabilian, Yunhong Wang and Liming Chen. *3-D Face Recognition Using eLBP-Based Facial Description and Local Feature Hybrid Matching*. IEEE Transactions on Information Forensics and Security, vol. 7, no. 5, pages 1551–1565, 2012. xvi, 50, 59, 60, 64, 68, 103, 104, 105, 113, 121, 137

[Hubel & Wiesel. 1962] D. H. Hubel and T. N. Wiesel. *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*. The Journal of Physiology, vol. 160, no. 1, pages 106–154, 1962. 110

[Hurdal *et al.* 2000] M.K. Hurdal, K. Stephenson, P. Bowers, D.W. Sumners and D.A. Rottenberg. *Coordinate systems for conformal cerebellar flat maps*. NeuroImage, vol. 11, no. 5; PART 2, pages 467–467, 2000. 148

[Huynh *et al.* 2013] Tri Huynh, Rui Min and Jean-Luc Dugelay. *An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D*

face data. In Computer Vision-ACCV 2012 Workshops, pages 133–145. Springer, 2013. 174

[I. Mpiperis & Strintzis 2008] S. Malassiotis I. Mpiperis and M. G. Strintzis. *Bilinear models for 3-d face and facial expression recognition*. IEEE Transactions on Information Forensics and Security, pages 498–511, 2008. 207, 208

[Ins 2011] Inspeck mega capturor ii digitizerurl. http://www.inspeck.com, 2011. 14

[J. Wang & Sun 2006] X. Wei J. Wang L.Yin and Y. Sun. *3d facial expression recognition based on primitive surface feature distribution*. In CVPR, pages 1399–1406, (2006. 207, 215, 216

[Jain *et al.* 1999] A.K. Jain, R. Bolle and S. Pankanti. Biometrics: Personal identification in networked society. Kluwer Academic Publications, 1999. 1

[Jin *et al.* 2008] M. Jin, J. Kim, F. Luo and X. Gu. *Discrete surface Ricci flow*. IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 5, pages 1030–1043, 2008. 144, 148, 150

[Johnson & Hebert 1999] Andrew E. Johnson and Martial Hebert. *Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes*. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 21, no. 5, pages 433–449, 1999. 62

[Kakadiaris *et al.* 2007a] Ioannis A. Kakadiaris, Georgios Passalis, George Toderici, Mohammed N. Murtuza, Yunliang Lu, Nikolaos Karampatziakis and Theoharis 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, pages 640–649, 2007. xiv, 30, 32, 33, 36, 37, 62, 65, 68, 70, 92, 94, 99, 103, 104, 105, 121, 128, 137, 173

[Kakadiaris *et al.* 2007b] Ioannis A. Kakadiaris, Georgios Passalis, George Toderici, Mohammed N. Murtuza, Yunliang Lu, Nikolaos Karampatziakis and Theoharis 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, pages 640–649, 2007. 251, 253, 254

[Kanade 1973] T. Kanade. *Picture processing by Computer Complex and Recognition of Human Face*. PhD thesis, Kyoto University, 1973. 8

[Kimmel & Sethian 1998] R. Kimmel and J. A. Sethian. *Computing Geodesic Paths on Manifolds*. In Proc. Natl. Acad. Sci. USA, pages 8431–8435, 1998. 44

[Kin 2011] Kinecturl. http://www.xbox.com/en-GB/kinect, 2011. 14

[Klare *et al.* 2011] B. Klare, A. A. Paulino and A. K. Jain. *Analysis of facial features in identical twins*. 2011. International Joint Conference on Biometrics (IJCB). 241, 242

[Klasing *et al.* 2009] K. Klasing, D. Althoff, D. Wollherr and M. Buss. *Comparison of surface normal estimation methods for range sensing applications*. In Proc. IEEE Int. Conf. on Robotics and Automation, 2009. 73

[L. Yin & Rosato. 2006] Y. Sun J. Wang L. Yin X. Wei and M. Rosato. *A 3D Facial Expression Database For Facial Behavior Research*. In The 7th International Conference on Automatic Face and Gesture Recognition, pages 211–216, (2006. 207, 213

[Lanckriet *et al.* 2004] G. Lanckriet, T. Bie, N. Cristianini, M. Jordan and W. Noble. *A statistical framework for genomic data fusion*. Bioinformatics, vol. 20, no. 16, pages 2626–2635, 2004. 218, 219

[Lee *et al.* 2003] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang and David Kriegman. *Video-based face recognition using probabilistic appearance manifolds*. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–313, 2003. 174

[Lee 1996] Tai Sing Lee. *Image Representation Using 2D Gabor Wavelets*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, pages 959–971, 1996. 245

[Lei *et al.* 2011] Zhen Lei, ShengCai Liao, Matti Pietikainen and Stan Z. Li. *Face Recognition by Exploring Information Jointly in Space, Scale and Orientation*. IEEE Transactions on Image Processing, vol. 20, no. 1, pages 247–256, 2011. 70

[Lemaire *et al.* 2011] Pierre Lemaire, Boulbaba Ben Amor, Mohsen Ardabilian, Liming Chen and Mohamed Daoudi. *Fully automatic 3D facial expression recognition using a region-based approach*. In Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding(J-HGBU), pages 53–58, 2011. 222, 223, 237, 238, 239

[Lévy & Bonneel 2013] Bruno Lévy and Nicolas Bonneel. *Variational Anisotropic Surface Meshing with Voronoi Parallel Linear Enumeration*. In Proceedings of the 21st International Meshing Roundtable, pages 349–366, 2013. 147

[Lévy *et al.* 2002] Bruno Lévy, Sylvain Petitjean, Nicolas Ray and Jérome Maillot. *Least squares conformal maps for automatic texture atlas generation*. In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '02), pages 362–371, 2002. 148

[Li & Jain 2005] S. Z. Li and A. K. Jain. Handbook of face recognition. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. xiii, 8, 9

[Li & Zhang 2007] X. Li and H. Zhang. *Adapting geometric attributes for expression-invariant 3D face recognition*. In IEEE International Conference on Shape Modeling and Applications, pages 21–32, Jun. 2007. 50, 61

[Li *et al.* 2009] X. Li, T. Jia and H. Zhang. *Expression-insensitive 3D face recognition using sparse representation*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 2575–2582, Jun. 2009. xvi, 50, 61, 62

[Li *et al.* 2011a] H. Li, D. Huang, J.M.Morvan and L. Chen. *Learning weighted sparse representation of encoded facial normal information for expression-robust 3D face recognition*. In Proc. IEEE Int. Joint Conf. on Biometrics, 2011. 71, 104, 106, 137

[Li *et al.* 2011b] H. Li, D. Huang, J.M.Morvan and L. Chen. *Learning weighted sparse representation of encoded facial normal information for expression-robust 3D face recognition*. In Proc. IEEE Int. Joint Conf. on Biometrics, 2011. 218, 219, 221, 242, 243, 244, 245, 250, 252

[Li *et al.* 2011c] Huibin Li, Di Huang, P. Lemaire, J-M Morvan and Liming Chen. *Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities*. In Proc. IEEE Int. Conf. on Image Processing, pages 3053–3056, 2011. 24, 65, 69, 101, 102, 104, 105, 109, 112, 113, 173

[Li *et al.* 2011d] Huibin Li, Jean-Marie Morvan and Liming Chen. *3D Facial Expression Recognition Based on Histograms of Surface Differential Quantities*. In Advances Concepts for Intelligent Vision Systems, volume 6915 of *Lecture Notes in Computer Science*, pages 483–494, 2011. 217, 218, 222, 223, 226, 237, 238

[Li *et al.* 2012] Huibin Li, Di Huang, Liming Chen, Yunhong Wang and J.-M. Morvan. *A group of facial normal descriptors for recognizing 3D identical twins*. In Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on, pages 271–277, 2012. 173

[Liu & Cheng 2003] Xiaoming Liu and Tsuhan Cheng. *Video-based face recognition using adaptive hidden markov models*. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–340, 2003. 174

[Liu *et al.* 2002] Q. Liu, R. Huang, H. Lu and S. Ma. *Face recognition using kernel-based fisher discriminant analysis*. In IEEE International Conference on Automatic Face and Gesture Recognition, pages 197–201, May 2002. 8

[Lo & Siebert 2009] Tsz-Wai Rachel Lo and J. Paul Siebert. *Local feature extraction and matching on range images: 2.5D SIFT*. Computer Vision and Image Understanding, vol. 113, no. 12, pages 1235–1250, 2009. 113

[Lowe 2004] David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*. Int. J. Comput. Vision, vol. 60, no. 2, pages 91–110, 2004. 60, 105, 112

[Lu & Jain 2005] X. Lu and A. K. Jain. *Deformation Analysis for 3D Face Matching*. In IEEE Workshop on Applications of Computer Vision, pages 99–104, 2005. 30

[Lu & Jain 2006] X. Lu and A. K. Jain. *Deformation modeling for robust 3D face matching*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1377–1383, 2006. 30

[Lu & Jain 2008] X. Lu and A. K. Jain. *Deformation Modeling for Robust 3D Face Matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 8, pages 1346–1357, 2008. xiv, 30, 64, 105

[Lu *et al.* 2004] X. Lu, D. Colbry and A. K. Jain. *Three-dimensional model based face recognition*. In International Conference on Pattern Recognition, pages 362–366, 2004. 30

[Lui *et al.* 2010] Lok Ming Lui, Tsz Wai Wong, Wei Zeng, Xianfeng Gu, Paul M. Thompson, Tony F. Chan and Shing Tung Yau. *Detection of Shape Deformities Using Yamabe Flow and Beltrami Coefficients*. Inverse Problems and Imaging, vol. 4, no. 2, pages 311–333, 2010. 148

[Maalej *et al.* 2010] A. Maalej, B. Ben Amor, M. Daoudi, A. Srivastava and S. Berretti. *Local 3D Shape Analysis for Facial Expression Recognition*. In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 4129 –4132, 2010. 226

[Maes *et al.* 2010] 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 Proc. Int. Conf. on Biometrics: Theory, Applications and Systems, 2010. 65, 109, 113, 128, 132, 133, 173

[Malassiotis & Tsalakanidou 2011] S. Malassiotis and F. Tsalakanidou. *Recognizing facial expressions from 3D video: Current results and future prospects*. In IEEE International Conference on Automatic Face Gesture Recognition and Workshops(FG), pages 597 –602, 2011. 226

[Marchandise *et al.* 2011] E. Marchandise, C. Carton de Wiart, W. G. Vos, C. Geuzaine and J.-F. Remacle. *High-quality surface remeshing using harmonic maps¡ª Part II: Surfaces with high genus and of large aspect ratio*. International Journal for Numerical Methods in Engineering, vol. 86, no. 11, pages 1303–1321, 2011. 144

[Marchandise *et al.* 2012] Emilie Marchandise, Jean-Fran?ois Remacle and Christophe Geuzaine. *Quality Surface Meshing Using Discrete Parametrizations*. In Proceedings of the 20th International Meshing Roundtable, pages 21–39, 2012. 25, 144

[Martinez 2002] A.M. Martinez. *Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 6, pages 748 –763, jun 2002. 70

[Maurer *et al.* 2005] Thomas Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West and G. Medioni. *Performance of Geometrix ActiveID TM 3D Face Recognition Engine on the FRGC Data*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 154–154, 2005. 104, 105

[Mei 2003] Xiangmin Mei. Differential geometry (in chinese). High Education Press, 2003. 175

[Metaxas & Kakadiaris 2002] D.N. Metaxas and I.A. Kakadiaris. *Elastically adaptive deformable models*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 10, pages 1310–1321, 2002. 33

[Meyer *et al.* 2001] ThomasH. Meyer, Marian Eriksson and RobertC. Maggio. *Gradient Estimation from Irregularly Spaced Data Sets*. Mathematical Geology, vol. 33, pages 693–717, 2001. 108

[Mian *et al.* 2006] Ajmal S Mian, Mohammed Bennamoun and Robyn Owens. *Three-dimensional model-based object recognition and segmentation in cluttered scenes*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, no. 10, pages 1584–1601, 2006. 50, 63, 65

[Mian *et al.* 2007] A. S. Mian, M. Bennamoun and R. A. Owens. *An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pages 1927–1943, 2007. 13, 50, 62, 63, 68, 92, 105, 106, 113, 137

[Mian *et al.* 2008] A. S. Mian, M. Bennamoun and R. A. Owens. *Keypoint Detection and Local Feature Matching for Textured 3D Face Recognition*. International Journal of Computer Vision, vol. 79, no. 1, pages 1–12, 2008. 63, 65, 68, 90, 92, 105, 113, 121, 133, 137

[Min 2011] Minolta vivid 910url. http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/features.html, 2011. 14

[Mohammadzade & Hatzinakos 2013] 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, pages 381 –397, 2013. 69, 95, 173

[Moreno & Sanchez 2004] A. B. Moreno and A. Sanchez. *GavabDB: A 3D Face Database*. In COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications, pages 77–82, 2004. 17

[Morvan & Thibert 2004] J. M. Morvan and B. Thibert. *Approximation of the normal vector field and the area of a smooth surface*. Discrete Computational Geometry, vol. 32, no. 3, pages 383–400, 2004. 25, 144, 145

[Morvan 2008a] Jean Marie Morvan. Generalized curvatures. Springer, 2008. 144, 145, 205, 231

[Morvan 2008b] J.M. Morvan. Generalized curvatures. Springer, 2008. 208, 209

[Mpiperis *et al.* 2007] I. Mpiperis, S. Malassiotis and M.G. Strintzis. *3-D Face Recognition With the Geodesic Polar Representation*. Information Forensics and Security, IEEE Transactions on, vol. 2, no. 3, pages 537–547, 2007. 45

[Mpiperis *et al.* 2008a] I. Mpiperis, S. Malassiotis and M. G. Strintzis. *Bilinear Models for 3-D Face and Facial Expression Recognition*. IEEE Transaction on IEEE Transactions on Information Forensics and Security, vol. 3, pages 498–511, 2008. xv, 30, 41, 42

[Mpiperis *et al.* 2008b] I. Mpiperis, S. Malassiotis and M.G. Strintzis. *Bilinear Models for 3-D Face and Facial Expression Recognition*. IEEE Transactions on Information Forensics and Security, vol. 3, no. 3, pages 498 –511, sept. 2008. 226, 237, 238, 239

[Nash 1954] John Nash. *C1 isometric imbeddings*. Annals of Mathematics, vol. 60, no. 3, pages 383–396, 1954. 164

[Nash 1956] John Nash. *The imbedding problem for Riemannian manifolds*. Annals of Mathematics, vol. 63, no. 1, pages 20–63, 1956. 164

[Ocegueda *et al.* 2011a] O. Ocegueda, G. Passalis, T. Theoharis, S.K. Shah and I.A. Kakadiaris. *UR3D-C: Linear Dimensionality Reduction for Efficient 3D Face Recognition*. In Proc. IEEE Int. Joint Conf. on Biometrics, 2011. 37, 92, 93, 99, 128, 132

[Ocegueda *et al.* 2011b] O. Ocegueda, S.K. Shah and I.A. Kakadiaris. *Which parts of the face give out your identity?* In IEEE Conference on Computer Vision and Pattern Recognition, pages 641–648, 2011. 37, 92, 93

[Ojala *et al.* 2002] 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, pages 971–987, 2002. 70, 84

[Otsuka *et al.* 2007] K. Otsuka, H. Sawada and J. Yamato. *Automatic inference of cross-modal nonverbal interactions in multiparty conversations: "who responds to whom, when, and how?" from gaze, head gestures, and utterances*. In ICMI, pages 255–262, 2007. 207

[P. Alliez 2003] O. Devillers B. Lévy M. Desbrun P. Alliez D. Cohen-Steiner. *Anisotropic polygonal remeshing*. ACM Trans. Graph., vol. 22, no. 3, pages 485–493, 2003. 208, 209

[Pantic & Rothkrantz 2000] Maja Pantic and Léon J. M. Rothkrantz. *Automatic Analysis of Facial Expressions: The State of the Art*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pages 1424–1445, 2000. 225

[Passalis *et al.* 2005] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici and N. Murtuza. *Evaluation of 3D Face Recognition in the presence of facial expressions: an Annotated Deformable Model approach*. In Computer Vision and Pattern Recognition, pages 171–171, 2005. xiv, 32, 33

[Passalis *et al.* 2011] G. Passalis, P. Perakis, T. Theoharis and I.A. Kakadiaris. *Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 10, pages 1938–1951, 2011. xiv, xv, 37, 38, 64, 65, 100, 104, 173

[Pati *et al.* 1993] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad. *Orthogonal Matching Pursuit: Recursive function approximation with applications to wavelet decomposition.* In Proc. 27th Asilomar Conf. on Signals, Systems and Computers, 1993. 78, 79, 84, 115, 233, 248

[Paul Chew 1989] L. Paul Chew. *Guaranteed-Quality Triangular Meshes*. In Department of Computer Science Tech Report 89-983, Cornell University, 1989. 147, 149, 152, 153

[Paul Chew 1993] L. Paul Chew. *Guaranteed-Quality Mesh Generation for Curved Surfaces*. In Proceedings of the Ninth Symposium on Computational Geometry, pages 274–280, 1993. 24, 144, 147, 149, 153

[Pears *et al.* 2010] Nick Pears, Tom Heseltine and Marcelo Romero. *From 3D Point Clouds to Pose-Normalised Depth Maps*. International Journal of Computer Vision, vol. 89, no. 2-3, pages 152–176, 2010. 104

[Pele & Werman 2010] O. Pele and M. Werman. *The Quadratic-Chi Histogram Distance Family*. In Proc. Int. Conf. European Conference on Computer Vision, 2010. 85

[Pentland & Choudhury 2000] A. Pentland and T. Choudhury. *Personalizing smart environments: Face recognition for human interaction*. IEEE Computer. Special issue on Biometrics, 2000. 4, 5

[Perakis *et al.* 2009] P. Perakis, G. Passalis, T. Theoharis, G. Toderici and I.A. Kakadiaris. *Partial matching of interpose 3D facial data for face recognition*. In IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, BTAS '09., pages 1–8, 2009. xiv, xv, 37, 38, 64, 173

[Phillips *et al.*] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. V. Bruegge, P. J. Grother, G. W. Quinn and M. Pruitt. *Distinguishing identical twins by face recognition*. IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2011. 241

[Phillips *et al.* 2000] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss. *The FERET evaluation methodology for face-recognition algorithms*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pages 1090–1104, Oct. 2000. 9, 98

[Phillips *et al.* 2005] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min and W. Worek. *Overview of the face recognition grand challenge*. In Proc. Int. Conf. Computer Vision and Pattern Recognition, 2005. 17, 67, 68, 80, 99, 116, 133

[Pike *et al.* 2000] G. Pike, R. Kemp and N. Brace. *The psychology of human face recognition*. In The psychology of human face recognition, pages 11/1–11/6, 2000. 6

[Pratikakis *et al.* 2010] I Pratikakis, M Spagnuolo, T Theoharis and R Veltkamp. *A robust 3D interest points detector based on Harris operator*. 2010. 65

[Queirolo *et al.* 2010] C.C. Queirolo, L. Silva, O.R.P. Bellon and M.P. Segundo. *3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, pages 206–219, 2010. xvi, 49, 54, 55, 56, 57, 69, 92, 94, 99, 137, 173

[Rakotomamonjy *et al.* 2008] Alain Rakotomamonjy, Francis R. Bach, Stèphane Canu and Yves Grandvalet. *SimpleMKL*. 2008. 219, 220, 221

[Remacle *et al.* 2010] J.-F. Remacle, C. Geuzaine, G. Comp¨¨re and E. Marchandise. *High-quality surface remeshing using harmonic maps*. International Journal for Numerical Methods in Engineering, vol. 83, no. 4, pages 403–425, 2010. 144

[Rosato *et al.* 2008] M. Rosato, Xiaochen Chen and Lijun Yin. *Automatic Registration of Vertex Correspondences for 3D Facial Expression Analysis*. In 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), pages 1–7, 2008. 226

[Ruggeri *et al.* 2010] Mauro R Ruggeri, Giuseppe Patanè, Michela Spagnuolo and Dietmar Saupe. *Spectral-driven isometry-invariant matching of 3D shapes*. International Journal of Computer Vision, vol. 89, no. 2-3, pages 248–265, 2010. 65

[Ruppert 1995] Jim Ruppert. *A Delaunay Refinement Algorithm for Quality 2-Dimensional Mesh Generation*. Journal of Algorithms, vol. 18, no. 3, pages 548–585, May 1995. 147, 149, 195

[Russ *et al.* 2006] T. Russ, C. Boehnen and T. Peters. *3D Face Recognition Using 3D Alignment for PCA*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1391–1398, 2006. 104

[Russ *et al.* 2009] T. Russ, C. Boehnen and T. Peters. *3D Face Recognition Using 3D Alignment for PCA*. In Computer Vision and Pattern Recognition, volume 2, pages 1391–1398, 2009. 105, 108

[S. Berretti 2010] P. Pala B. Ben Amor M. Daoudi S. Berretti A. D. Bimbo. *A Set of Selected SIFT Features for 3D Facial Expression Recognition*. In ICPR, pages 4125–4128, 2010. 207, 208, 216

[Samal & Iyengar 1992] Ashok Samal and Prasana A. Iyengar. *Automatic recognition and analysis of human faces and facial expressions: a survey*. Pattern Recognition, vol. 25, no. 1, pages 65 – 77, 1992. 225

[Samir *et al.* 2006] 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, pages 1858–1863, 2006. xv, 45, 46, 64, 65, 68, 104

[Samir *et al.* 2009] Chafik Samir, Anuj Srivastava, Mohamed Daoudi and Eric Klassen. *An Intrinsic Framework for Analysis of Facial Surfaces*. International Journal of Computer Vision, vol. 82, no. 1, pages 80–95, 2009. xv, 45, 46, 64

[Sandbach *et al.* 2012] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic and Lijun Yin. *Static and dynamic 3D facial expression recognition: A comprehensive survey*. Image and Vision Computing, 2012. in press. 226

[Savran & Sankur 2009] A. Savran and B. Sankur. *Automatic detection of facial actions from 3D data*. In IEEE 12th International Conference on Computer Vision Workshops (ICCV), pages 1993–2000, 2009. 226

[Savran *et al.* 2008] A. Savran, N. Alyüz, H. Dibeklioglu, Oya Çeliktutan, B. Gökberk, B. Sankur and L. Akarun. *3D Face Recognition Benchmarks on the Bosphorus Database with Focus on Facial Expressions*. In Proc. Workshop on Biometrics and Identity Management, 2008. 17, 80, 98, 102, 116, 118

[Schoen & Yau 1994] Richard Schoen and Shing-Tung Yau. Lectures on differential geometry. International Press of Boston, 1994. 228

[Segundo *et al.* 2007] M. P. Segundo, C. Queirolo, O. R. P. Bellon and L. Silva. *Automatic 3D facial segmentation and landmark detection*. In International Conference on Image Analysis and Processing, 2007. 54

[Seitz *et al.* 2006] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein and Richard Szeliski. *A comparison and evaluation of multi-view stereo reconstruction algorithms*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 519–528, 2006. 14

[Sha *et al.* 2011] Teng Sha, Mingli Song, Jiajun Bu, Chun Chen and Dacheng Tao. *Feature level analysis for 3D facial expression recognition*. Neurocomputing, vol. 74, no. 12-13, pages 2135 – 2141, 2011. 226

[Sheffer & de Sturler 2001] Alla Sheffer and Eric de Sturler. *Parameterization of Faceted Surfaces for Meshing using Angle-Based Flattening*. Engineering With Computers, vol. 17, no. 3, pages 326–337, 2001. 148

[Sheffer *et al.* 2006] Alla Sheffer, Emil Praun and Kenneth Rose. *Mesh parameterization methods and their applications*. Foundations and Trends in Computer Graphics and Vision, vol. 2, no. 2, pages 105–171, 2006. 148

[Shewchuk 2001] Jonathan Richard Shewchuk. *Delaunay Refinement Algorithms for Triangular Mesh Generation*. Computational Geometry: Theory and Applications, vol. 22, pages 1–3, 2001. 147

[Shifrin 2010] Theodore Shifrin. Differential geometry: A first course in curves and surfaces. Preliminary Version, 2010. 175

[Silva *et al.* 2005] Luciano Silva, Olga R. P. Bellon and Kim L. Boyer. *Precision Range Image Registration Using a Robust Surface Interpenetration Measure and Enhanced Genetic Algorithms*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pages 762–776, 2005. xvi, 54, 55

[Sinha *et al.* 2006] Pawan Sinha, Benjamin J Balas, Yuri Ostrovsky and Richard Russell. *Face recognition by humans*. Face Processing: Advanced Modeling and Methods. Academic Press, San Diego, 2006. xiii, 6, 7, 13

[Smeets *et al.* 2009] D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen and P. Suetens. *Isometric deformation modeling using singular value decomposition for 3D expression-invariant face recognition*. In Proc. Int. Conf. on Biometrics: Theory, Applications, and Systems, pages 1–6, 2009. 105

[Smeets *et al.* 2012] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen and P. Suetens. *A Comparative Study of 3-D Face Recognition Under Expression Variations*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 42, no. 5, pages 710–727, 2012. 103

[Smeets *et al.* 2013] Dirk Smeets, Johannes Keustermans, Dirk Vandermeulen and Paul Suetens. *meshSIFT: Local surface features for 3D face recognition under expression variations and partial data*. Computer Vision and Image Understanding, vol. 117, no. 2, pages 158–169, 2013. 92, 104, 113, 136, 137

[Soyel & Demirel. 2007a] H. Soyel and H. Demirel. *Facial expression recognition using 3d facial feature distances*. In In Int. Conf. on Image Analysis and Recognition, pages 831–838, (2007. 207, 216

[Soyel & Demirel 2007b] Hamit Soyel and Hasan Demirel. *Facial Expression Recognition Using 3D Facial Feature Distances*. In Image Analysis and Recognition, volume 4633 of *Lecture Notes in Computer Science*, pages 831–838, 2007. 217, 218, 222, 223, 226, 237, 238, 239

[Soyel & Demirel 2008] H. Soyel and H. Demirel. *3D facial expression recognition with geometrically localized facial features*. In 23rd International Symposium on Computer and Information Sciences (ISCIS), pages 1 –4, 2008. 226, 237, 238

[Spreeuwers 2011] L.J. Spreeuwers. *Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region classifiers.* International Journal of Computer Vision, vol. 93, no. 3, pages 389–414, 2011. xvi, 49, 56, 69, 92, 94, 95, 99, 103, 104, 134, 137

[Springborn *et al.* 2008] Boris Springborn, Peter Schröder and Ulrich Pinkall. *Conformal equivalence of triangle meshes*. ACM Trans. Graph., vol. 27, no. 3, pages 77:1–77:11, August 2008. 148

[Sun & Yin 2008] Y. Sun and L. Yin. *Facial Expression Recognition Based on 3D Dynamic Range Model Sequences*. In 10th European Conference on Computer Vision, pages 58–71, 2008. 207

[Sun *et al.* 2009] Jian Sun, Maks Ovsjanikov and Leonidas Guibas. *A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion*. Computer Graphics Forum, vol. 28, no. 5, pages 1383–1392, 2009. 65, 142, 144

[Sun *et al.* 2010] Z. Sun, A. A. Paulino, J. Feng, Z. Chai, T. Tan and A. K. Jain. *A study of multibiometric traits of identical twins*. In SPIE Biometric technology for human identification, 2010. 241

[Sun *et al.* 2011] D. Sun, W. Sung and R. Chen. *3D Face Recognition Based on Local Curvature Feature Matching*. Applied Mechanics and Materials, vol. 121, no. 126, pages 609–616, 2011. 69

[Surazhsky *et al.* 2003] Vitaly Surazhsky, Pierre Alliez and Craig Gotsman. *Isotropic remeshing of surfaces: A local parameterization approach*. In Proceedings of 12th International Meshing Roundtable, pages 215–224, 2003. 147

[Szeptycki *et al.* 2009a] P. Szeptycki, M. Ardabilian and L. Chen. *A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking*. In International Conference on Biometrics: Theory, Applications and Systems (BTAS), pages 3206–3211, 2009. 68, 81, 119

[Szeptycki *et al.* 2009b] P. Szeptycki, M. Ardabilian and L. Chen. *A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking*. In International Conference on Biometrics: Theory, Applications and Systems (BTAS), pages 3206–3211, 2009. 249

[Szeptycki *et al.* 2009c] Przemyslaw Szeptycki, Mohsen Ardabilian and Liming Chen. *A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking*. In Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems, (BTAS), pages 32–37, 2009. 226, 227, 230

[Szeptycki *et al.* 2010] P. Szeptycki, M. Ardabilian, L. Chen, W. Zeng, X. Gu and D. Samaras. *Conformal Mapping-based 3D Face Recognition*. In 3D Data Processing, Visualization and Transmission Symposium (3DPVT'10), Paris, France, May 17-20 2010. 226

[Tanaka & Ikeda 1996] H. T. Tanaka and M. Ikeda. *Curvature-based face surface recognition using spherical correlation - principal directions for curved object recognition*. In International Conference on Pattern Recognition, volume 3, pages 638–642, Aug. 1996. 50, 61

[Tanaka *et al.* 1998] H. T. Tanaka, M. Ikeda and H. Chiaki. *Curvature-based face surface recognition using spherical correlation. Principal directions for curved object recognition*. In IEEE International Conference on Automatic Face and Gesture Recognition, pages 372–377, Apr. 1998. xvi, 50, 61

[Tang & Huang 2008a] H. Tang and T. S. Huang. *3d facial expression recognition based on automatically selected features*. In In Int. Conf. on Computer Vision and Pattern Recognition, pages 1–8, 2008. 207, 216

[Tang & Huang 2008b] Hao Tang and T.S. Huang. *3D facial expression recognition based on automatically selected features*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), pages 1–8, 2008. 217, 218, 222, 223, 226, 237, 238, 239

[Tang & Huang 2008c] Hao Tang and T.S. Huang. *3D facial expression recognition based on properties of line segments connecting facial feature points*. In 8th IEEE International Conference on Automatic Face Gesture Recognition(FG), pages 1–6, 2008. 226, 237, 238

[Tang *et al.* 2013] Hengliang Tang, Baocai Yin, Yanfeng Sun and Yongli Hu. *3D face recognition using local binary patterns*. Signal Processing, vol. 93, no. 8, pages 2190–2198, 2013. 68, 92

[Taubin 1995] G. Taubin. *Estimating the tensor of curvature of a surface from a polyhedral approximation*. In ICCV, 1995. 208

[Tenenbaum & Freeman 2000] J. Tenenbaum and W. Freeman. *Separating style and content with bilinear models*. Neural Computation, vol. 12, page 1247¨C1283, 2000. 41, 42

[Tenenbaum *et al.* 2000] J. B. Tenenbaum, V. Silva and J. C. Langford. *A global geometric framework for nonlinear dimensionality reduction*. Science, vol. 290, no. 5500, pages 2319–2323, 2000. 9

[ter Haar *et al.* 2008] F.B. ter Haar, M. Daoudi and R.C. Veltkamp. *SHape REtrieval contest 2008: 3D face scans*. In IEEE International Conference on Shape Modeling and Applications,, pages 225–226, 2008. 48

[ter Haara & Veltkampb 2010] Frank B. ter Haara and Remco C. Veltkampb. *Expression modeling for expression-invariant face recognition*. Computers Graphics, vol. 34, no. 3, pages 231 – 241, 2010. 41

[Terzopoulos & Metaxas 1991] D. Terzopoulos and D. Metaxas. *Dynamic 3D models with local and global deformations: deformable superquadrics*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 13, no. 7, pages 703–714, 1991. 33, 34, 36

[Tombari *et al.* 2013] Federico Tombari, Samuele Salti and Luigi Di Stefano. *Performance evaluation of 3D keypoint detectors*. International Journal of Computer Vision, pages 1–23, 2013. 65

[Torr & Zisserman 2000] P. H. S. Torr and A. Zisserman. *MLESAC: A New Robust Estimator with Application to Estimating Image Geometry*. Computer Vision and Image Understanding, vol. 78, page 2000, 2000. 54

[Tsalakanidou *et al.* 2005] Filareti Tsalakanidou, Frank Forster, Sotiris Malassiotis and Michael G. Strintzis. *Real-time acquisition of depth and color images using structured light and its application to 3D face recognition*. Real-Time Imaging, vol. 11, no. 5-6, pages 358–369, 2005. xiii, 14

[Turk & Pentland 1991] M. Turk and A. Pentland. *Eigenfaces for face recognition*. Journal of Cognitive Neuroscience, vol. 3, no. 1, pages 71–86, 1991. 8

[Twi] *Twins days*. Available: http://www.twinsdays.org/. 241, 248

[UND 2008] University of notre dame biometrics database. http://www.nd.edu/¡«cvrl/UNDBiometricsDatabase.html, 2008. 17

[Veltkamp *et al.* 2011] Remco C. Veltkamp, Stefan van Jole, Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, Huibin Li, Liming Chen, Peter Claes, Dirk Smeets, Jeroen Hermans, Dirk Vandermeulen and Paul Suetens. *SHREC '11 Track: 3D Face Models Retrieval*. In 3DOR, pages 89–95, 2011. 17, 102, 105

[Venkatesh *et al.* 2009] Y. V. Venkatesh, Ashraf A. Kassim and O. V. Ramana Murthy. *A novel approach to classification of facial expressions from 3D-mesh datasets using modified PCA*. Pattern Recogn. Lett., vol. 30, pages 1128–1137, 2009. 226

[Vijayan *et al.* 2011a] V. Vijayan, K. W. Bowyer and P. J. Flynn. *3D Twins and Expression Challenge*. In Proc. Int. Conf. on Computer Vision Workshops, 2011. 17, 69, 80, 241, 242

[Vijayan *et al.* 2011b] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen, O. Ocegueda, S. K. Shah and I. A. Kakadiaris. *Twins 3D face recognition challenge*. In Proc. Int. Joint Conf. on Biometrics, 2011. ix, xi, 69, 70, 80, 81, 83, 84, 102, 133, 241, 242, 248, 249, 251, 252

[Wagner *et al.* 2012] Andrew Wagner, John Wright, Arvind Ganesh, Zihan Zhou, Hossein Mobahi and Yi Ma. *Towards a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 34, no. 2, pages 372–386, 2012. 232

[Wang *et al.* 2005] Yalin Wang, Xianfeng Gu, Kiralee M. Hayashi, Tony F. Chan, Paul M. Thompson and Shing-Tung Yau. *Surface Parameterization Using Riemann Surface Structure*. In International Conference on Computer Vision, pages 1061–1066, 2005. 148

[Wang *et al.* 2006a] Jun Wang, Lijun Yin, Xiaozhou Wei and Yi Sun. *3D Facial Expression Recognition Based on Primitive Surface Feature Distribution*. In Computer Vision and Pattern Recognition, volume 2, pages 1399–1406, 2006. 107

[Wang *et al.* 2006b] Jun Wang, Lijun Yin, Xiaozhou Wei and Yi Sun. *3D Facial Expression Recognition Based on Primitive Surface Feature Distribution*. In IEEE Conference on Computer Vision and Pattern(CVPR), pages 1399–1406, 2006. 217, 218, 222, 223, 226, 237, 238, 239

[Wang *et al.* 2006c] Yueming Wang, Gang Pan, Zhaohui Wu and Yigang Wang. *Exploring Facial Expression Effects in 3D Face Recognition Using Partial ICP*. In 7th Asian Conference on Computer Vision, volume 3851, pages 581–590, 2006. 104, 105

[Wang *et al.* 2007] Sen Wang, Yang Wang, Miao Jin, Xianfeng David Gu and Dimitris Samaras. *Conformal Geometry and Its Applications on 3D Shape Matching, Recognition, and Stitching*. IEEE TPAMI, vol. 29, no. 7, pages 1209 – 1220, 2007. 226

[Wang *et al.* 2008] Yang Wang, Mohit Gupta, Song Zhang, Sen Wang, Xianfeng Gu, Dimitris Samaras and Peisen Huang. *High Resolution Tracking of Non-Rigid Motion of Densely Sampled 3D Data Using Harmonic Maps*. International Journal of Computer Vision, vol. 76, no. 3, pages 283–300, 2008. 144

[Wang *et al.* 2009] Huafeng Wang, Yunhong Wang and Yuan Cao. *Video-based face recognition: A survey*. World Academy of Science, Engineering and Technology, vol. 60, pages 293–302, 2009. 174

[Wang *et al.* 2010] 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, pages 1858–1870, 2010. xvi, 50, 58, 59, 65, 68, 99, 103, 134, 137

[Weyrauch *et al.* 2004] B. Weyrauch, B. Heisele, J. Huang and V. Blanz. *Component-Based Face Recognition with 3D Morphable Models*. In Computer Vision and Pattern Recognition Workshop, pages 85–85, 2004. 106

[Wilmer *et al.* 2010] Jeremy B. Wilmer, Laura Germine, Christopher F. Chabris, Garga Chatterjee, Mark Williams, Eric Loken, Ken Nakayama and Bradley Duchaine. *Human face recognition ability is specific and highly heritable*. Proceedings of the National Academy of Sciences, vol. 107, no. 11, pages 5238–5241, 2010. 6

[Wiskott *et al.* 1997a] L. Wiskott, J.-M. Fellous, N. Kuiger and C. von der Malsburg. *Face recognition by elastic bunch graph matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 775–779, Jul. 1997. 9

[Wiskott *et al.* 1997b] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger and Christopher von der Malsburg. *Face Recognition by Elastic Bunch Graph Matching*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pages 775–779, 1997. 243

[Woodham 1980] Robert J Woodham. *Photometric method for determining surface orientation from multiple images*. Optical Engineering, vol. 19, no. 1, pages 191139–191139–, 1980. 15

[Wright *et al.* 2009a] 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, pages 210–227, 2009. 24, 101, 112, 114

[Wright *et al.* 2009b] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry and Yi Ma. *Robust Face Recognition via Sparse Representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 31, no. 2, pages 210 –227, feb. 2009. 226, 232

[Wright *et al.* 2009c] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry and Yi Ma. *Robust Face Recognition via Sparse Representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pages 210–227, 2009. 248, 251

[Wu *et al.* 2004] Z. Wu, Y. Wang and G. Pan. *3D face recognition using local shape map*. In IEEE International Conference on Image Processing, volume 3, pages 2003–2006, Oct. 2004. 50, 62

[X. Zhao 2009] E. Dellandrea L. Chen X. Zhao P. Szeptycki. *Precise 2.5D Facial Landmarking via an Analysis by Synthesis approach*. In 2009 IEEE Workshop on Applications of Computer Vision, 2009. 216

[Xin *et al.* 2011] Shi Qing Xin, Shuang Min Chen, Ying He, Guo Jin Wang, Xianfeng Gu and Hong Qin. *Isotropic Mesh Simplification by Evolving the Geodesic Delaunay Triangulation*. In Proceedings of the Eighth International Symposium on Voronoi Diagrams in Science and Engineering (ISVD '11), pages 39–47, 2011. 145

[Xu *et al.* 2009] C. Xu, S. Li, T. Tan and L. Quan. *Automatic 3D face recognition from depth and intensity Gabor features*. Pattern Recognition, vol. 42, no. 9, pages 1895–1905, 2009. 243, 245, 246

[Yan *et al.* 2009] Dongming Yan, Bruno Lévy, Yang Liu, Feng Sun and Wenping Wang. *Isotropic Remeshing with Fast and Exact Computation of Restricted Voronoi Diagram*. In Proceedings of the Symposium on Geometry Processing, pages 1445–1454, 2009. 147

[Yang 2002] M.-H. Yang. *Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods*. In IEEE International Conference on Automatic Face and Gesture Recognition, pages 215–220, may 2002. 8

[Yin *et al.* 2006a] L. Yin, X. Wei, Y. Sun, J. Wang and M. Rosato. *A 3D Facial Expression Database For Facial Behavior Research*. In Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition, 2006. 17, 80

[Yin *et al.* 2006b] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato. *A 3D Facial Expression Database For Facial Behavior Research*. IEEE International Conference on Automatic Face and Gesture Recognition (FG), pages 211–216, 2006. 217, 221, 233

[Yin *et al.* 2008a] L. Yin, X. Chen, Y. Sun, T. Worm and M. Reale. *A High-Resolution 3D Dynamic Facial Expression Database*. In IEEE Int. Conference on Automatic Face Gesture Recognition, pages 1–6, 2008. 207

[Yin *et al.* 2008b] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm and Michael Reale. *A high-resolution 3D dynamic facial expression database*. In 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG'08., pages 1–6, 2008. 14, 174

[Ying *et al.* 2010] Zi-Lu Ying, Zhe-Wei Wang and Ming-Wei Huang. *Facial Expression Recognition Based on Fusion of Sparse Representation*. In Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, volume 6216 of *Lecture Notes in Computer Science*, pages 457–464. 2010. 226

[Yu *et al.* 2013] Tsz-Ho Yu, Oliver Woodford and Roberto Cipolla. *A Performance Evaluation of Volumetric 3D Interest Point Detectors*. International Journal of Computer Vision, vol. 102, no. 1-3, pages 180–197, 2013. 142

[Zafeiriou & Petrou 2010] S. Zafeiriou and M. Petrou. *Sparse representations for facial expressions recognition via l1 optimization*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPR), pages 32 –39, june 2010. 226

[Zafeiriou *et al.* 2011] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith and L. Smith. *The photoface database*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Workshop on Biometrics) (CVPRW2011), 2011. 15

[Zaharescu *et al.* 2009] Andrei Zaharescu, Edmond Boyer, Kiran Varanasi and Radu Horaud. *Surface feature detection and description with applications to mesh matching*. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 373–380. IEEE, 2009. 65, 109

[Zeng *et al.* 2008] W. Zeng, X. T. Yin, Y. Zeng, Y. K. Lai, X. Gu and D. Samaras. *3D Face Matching and Registration Based on Hyperbolic Ricci Flow*. In The CVPR Workshop on 3D Face Processing (CVPR'08-3DFP), 2008. 226

[Zeng *et al.* 2009] Zhihong Zeng, Maja Pantic, Glenn I. Roisman and Thomas S. Huang. *A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pages 39–58, 2009. 225

[Zeng *et al.* 2010a] Wei Zeng, Dimitris Samaras and Xianfeng David Gu. *Ricci Flow for 3D Shape Analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pages 662–677, 2010. 148, 230

[Zeng *et al.* 2010b] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, Dimitris Samaras and Nikos Paragios. *Dense non-rigid surface registration using high-order graph matching*. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 382–389. IEEE, 2010. 174

[Zeng *et al.* 2011] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, Dimitris Samaras and Nikos Paragios. *Intrinsic dense 3D surface tracking*. In CVPR, pages 1225–1232, 2011. 174

[Zhang & Huang 2004] Song Zhang and Peisen Huang. *High-Resolution, Real-time 3D Shape Acquisition*. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, CVPRW '04, pages 28–, 2004. 14

[Zhang & Wang 2009] Guangpeng Zhang and Yunhong Wang. *Faceprint: Fusion of Local Features for 3D Face Recognition*. In Advances in Biometrics, volume 5558 of *Lecture Notes in Computer Science*, pages 394–403. Springer Berlin Heidelberg, 2009. 113

[Zhang & Wang 2011] G. Zhang and Y. Wang. *Robust 3D face recognition based on resolution invariant features*. Pattern Recognition Letters, vol. 32, no. 7, pages 1009–1019, 2011. 69

[Zhang & Yau 2006] Song Zhang and Shing-Tung Yau. *High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method*. Optics Express, vol. 14, no. 7, pages 2644–2649, 2006. 14

[Zhang *et al.* 2005] W. Zhang, S. Shan, W. Gao, X. Chen and H. Zhang. *Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition*. In IEEE International Conference on Computer Vision (ICCV), pages 786–791, 2005. 243, 246

[Zhang *et al.* 2012] Haichao Zhang, Nasser M. Nasrabadi, Yanning Zhang and Thomas S. Huang. *Joint dynamic sparse representation for multi-view face recognition*. Pattern Recognition, vol. 45, no. 4, pages 1290–1298, 2012. 10, 115

[Zhang 1994] Z.Y. Zhang. *Iterative point matching for registration of free-form curves and surfaces*. International Journal of Computer Vision, vol. 13, no. 2, pages 119–152, 1994. 17, 82, 121

[Zhao *et al.* 2003] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld. *Face recognition: A literature survey*. ACM Computing Survey, vol. 35, pages 399–458, Dec. 2003. 8, 9, 67, 98, 105

[Zhao *et al.* 2010] Xi Zhao, Di Huang, Emmanuel Dellandrea and Liming Chen. *Automatic 3D Facial Expression Recognition Based on a Bayesian Belief Net and a Statistical Facial Feature Model*. International Conference on Pattern Recognition (ICPR), pages 3724–3727, 2010. 226, 237, 238, 239

[Zhao *et al.* 2011] Xi Zhao, Emmanuel Dellandréa, Liming Chen and Ioannis A. Kakadiaris. *Accurate Landmarking of Three-Dimensional Facial Data in the Presence of Facial Expressions and Occlusions Using a Three-Dimensional Statistical Facial Feature Model*. IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 41, no. 5, pages 1417–1428, 2011. 68, 104