Contributions à la biométrie: courbures, reconnaissance du visage sur résolutions transversales hétérologues et anti-spoofing
Yinhang Tang

To cite this version:


HAL Id: tel-01533439
https://tel.archives-ouvertes.fr/tel-01533439
Submitted on 6 Jun 2017

HAL est un dépôt open access multi-disciplinaire, ouvert à l’ensemble des chercheurs. Il est destiné à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
contributions to biometrics: curvatures, heterogeneous cross-resolution FR and anti-spoofing

YINHANG TANG
December 2016

JURY

<table>
<thead>
<tr>
<th>Prof. Alice CAPLIER</th>
<th>Grenoble-INP</th>
<th>Présidente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof. Jean-Luc DUGELAY</td>
<td>Eurecom</td>
<td>Rapporteur</td>
</tr>
<tr>
<td>Prof. Boulbaba BEN AMOR</td>
<td>Institut Mines-Télécom/TELECOM Lille</td>
<td>Rapporteur</td>
</tr>
<tr>
<td>Prof. Jean-Marie MORVAN</td>
<td>Université Lyon 1 &amp; KAUST</td>
<td>Directeur de thèse</td>
</tr>
<tr>
<td>Prof. Liming CHEN</td>
<td>Ecole Centrale de Lyon</td>
<td>Co-directeur de thèse</td>
</tr>
</tbody>
</table>
Acknowledgement

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

First and foremost, I would like to express my deeply felt thanks to my supervisors, Prof. Jean-Marie Morvan and Prof. Liming Chen, for their continuous support and encouragement, and also for their direction and countless hours spent with me on the research and this thesis. Working under the supervision of such knowledgeable, understanding and easy-going persons has been a great pleasure for me.

I would also like to express my great gratitude to each of my committee members: Prof. Alice Caplier, Prof. Jean-Luc Dugelay and Prof. Boulbaba Ben Amor, for their time, interest and helpful comments.

I express my special thanks to Dr. Xiang Sun and Dr. Huibin Li for their luminous guidance and helpful suggestions in geometry differential and 3D face recognition research, and for all those unforgettable moments when we worked together. 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. Di Huang, Dr. Boyang Gao, Dr. Xiaofang Wang, Dr. Yuxing Tang, Dr. Huaxiong Ding, Wuming Zhang, Ying Lv, Dr. Chen Wang, Fei Zheng, Zehua Fu. I am also grateful to my friends at the Ecole Centrale de Lyon, Dr. Xu Chen, Dr. Lu Zhang, Dr. Xingrong Huang, Li Wang, Yang Xu, Tao Zheng, for all the impressive time we spend together.

In special, I would like to express my deepest gratitude and love to my family. My father, Dingxian Tang, always be the greatest person in my heart. He gives me the opportunity to receive the higher education and changes my life. My mother, Rong Liu, always gives me her deepest love, and shows her support for the countless years of my academic career. I particularly acknowledge my wife, Xing Liu, 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. Especially, during the last two months before the defence of thesis, she stayed with me showing her patience, support and kindness.

Finally, 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.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>xv</td>
</tr>
<tr>
<td>Résumé</td>
<td>xix</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Context and Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Biometrics: An Innovation in Authentication</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Face: First Choice of Biometric Trait</td>
<td>5</td>
</tr>
<tr>
<td>1.1.3 Overview of Face Recognition</td>
<td>15</td>
</tr>
<tr>
<td>1.1.4 Brief Review in 2D Face Recognition</td>
<td>7</td>
</tr>
<tr>
<td>1.1.5 Challenging Issues in 2D Face Recognition</td>
<td>10</td>
</tr>
<tr>
<td>1.1.6 Opportunities for 3D Face Recognition</td>
<td>14</td>
</tr>
<tr>
<td>1.2 3D Face Recognition</td>
<td>15</td>
</tr>
<tr>
<td>1.2.1 3D Face Acquisition Techniques and Databases</td>
<td>15</td>
</tr>
<tr>
<td>1.2.2 Basic Concepts and Terminology</td>
<td>17</td>
</tr>
<tr>
<td>1.2.3 Main Challenges</td>
<td>20</td>
</tr>
<tr>
<td>1.3 Thesis Contributions</td>
<td>23</td>
</tr>
<tr>
<td>1.3.1 Methodologies</td>
<td>23</td>
</tr>
<tr>
<td>1.3.2 Main Contributions</td>
<td>24</td>
</tr>
<tr>
<td>1.4 Thesis Organization</td>
<td>27</td>
</tr>
<tr>
<td>2 Literature Review: 3D Face Recognition</td>
<td>29</td>
</tr>
<tr>
<td>2.1 Holistic Feature based Approaches</td>
<td>30</td>
</tr>
<tr>
<td>2.1.1 Subspace based approaches (PCA and LDA)</td>
<td>30</td>
</tr>
<tr>
<td>2.1.2 Iterative Closest Point (ICP)</td>
<td>32</td>
</tr>
<tr>
<td>2.1.3 Deformable Model</td>
<td>34</td>
</tr>
<tr>
<td>2.1.4 Iso-level curve based approaches</td>
<td>36</td>
</tr>
<tr>
<td>2.2 Local Region-wise Feature based Approaches</td>
<td>38</td>
</tr>
<tr>
<td>2.3 Local Point-wise Feature based Approaches</td>
<td>43</td>
</tr>
<tr>
<td>2.3.1 Geometry-texture descriptor based approaches</td>
<td>44</td>
</tr>
<tr>
<td>2.3.2 Geometry-shape descriptor based approaches</td>
<td>47</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>52</td>
</tr>
<tr>
<td>2.5 Summary</td>
<td>53</td>
</tr>
<tr>
<td>3 Reminders on Principal Curvature Measures</td>
<td>55</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>55</td>
</tr>
<tr>
<td>3.2 Background on currents</td>
<td>56</td>
</tr>
</tbody>
</table>
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1</td>
<td>General currents</td>
<td>56</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Rectifiable currents</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>The normal cycle</td>
<td>57</td>
</tr>
<tr>
<td>3.4</td>
<td>Curvature functions, curvature measures for smooth surfaces</td>
<td>58</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Classical principal curvatures of smooth surfaces in $\mathbb{E}^3$</td>
<td>59</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Second fundamental measures on smooth surfaces of smooth surfaces in $\mathbb{E}^3$</td>
<td>59</td>
</tr>
<tr>
<td>3.5</td>
<td>Curvature measures for singular spaces</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Principal curvature measures of triangle meshes in $\mathbb{E}^3$</td>
<td>62</td>
</tr>
<tr>
<td>3.7</td>
<td>A convergence theorem</td>
<td>62</td>
</tr>
<tr>
<td>3.8</td>
<td>Conclusion</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>3D Face Recognition with Curvature Faces based on Principal Curvature Measures</td>
<td>67</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>Local Principal Curvature Measures Pattern (LPCMP) feature descriptor</td>
<td>68</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Generating Curvature Faces</td>
<td>69</td>
</tr>
<tr>
<td>4.2.2</td>
<td>LPCMP feature descriptor extracted in curvature faces</td>
<td>71</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Sparse Representation-based Classifier and Fusion</td>
<td>72</td>
</tr>
<tr>
<td>4.3</td>
<td>Experiments</td>
<td>73</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Database</td>
<td>73</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Experiment Settings</td>
<td>74</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Experiment Results</td>
<td>75</td>
</tr>
<tr>
<td>4.3.3.1</td>
<td>Impact of region size of Borel subset</td>
<td>75</td>
</tr>
<tr>
<td>4.3.3.2</td>
<td>The impact of encoding pattern</td>
<td>76</td>
</tr>
<tr>
<td>4.3.3.3</td>
<td>Identification test with score-level fusion rules</td>
<td>77</td>
</tr>
<tr>
<td>4.3.3.4</td>
<td>Verification experiments</td>
<td>79</td>
</tr>
<tr>
<td>4.3.3.5</td>
<td>Comparison with the-state-of-the-art</td>
<td>80</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusion</td>
<td>82</td>
</tr>
<tr>
<td>5</td>
<td>3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations</td>
<td>83</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>83</td>
</tr>
<tr>
<td>5.2</td>
<td>Principal Curvature Measures based 3D Face Description and Recognition</td>
<td>86</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Framework of PCM-meshSIFT based 3D face recognition</td>
<td>86</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Principal curvature measures based 3D keypoint detection</td>
<td>87</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Principal curvature measures based 3D keypoint description</td>
<td>89</td>
</tr>
<tr>
<td>5.2.3.1</td>
<td>Canonical directions assignment</td>
<td>89</td>
</tr>
</tbody>
</table>
## 5.2.3.2 3D keypoint descriptor representation

5.2.4 SRC-based 3D keypoint matching and score level fusion

5.2.5 Experiments

5.2.5.1 Databases

5.2.5.2 Experimental settings

5.2.5.3 3D face recognition on FRGCv2 database

5.2.5.4 3D face recognition on Bosphorus database

5.2.5.5 Times cost

5.3 Heterogeneous Cross-Resolution related Experiments

5.3.1 Lock3DFace database

5.3.2 Signed Distance Function based facial surface optimization

5.3.3 Landmarks location with modified Active Shape Models

5.3.4 Experiments

5.3.4.1 Experiment Settings

5.3.4.2 3D face recognition with homogeneous resolution data

5.3.4.3 3D face recognition with heterogeneous resolutions

5.3.4.4 Time Cost

5.4 Conclusion

6 Shape Analysis based Anti-spoofing 3D Face Recognition with Mask Attacks

6.1 Introduction

6.2 Related Work

6.3 Experiments

6.3.1 Database

6.3.2 Experiment Scenarios

6.3.3 Experimental Results and Analysis

6.3.4 Comparison with the state-of-the-art approaches

6.4 Conclusion

7 Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

7.1 Introduction

7.1.1 Overview of Hand-dorsa Vein Recognition

7.2 Hand-dorsa vein image acquisition

7.3 Framework of multi-source keypoint based hand-dorsa vein recognition

7.4 Multi-level Keypoint Detection

7.4.1 The Difference of Gaussian Detector

7.4.2 Harris and Hessian Keypoint Detection
Contents

7.4.3 Design of Multi-level Keypoint Detection ............................................. 139
7.5 Oriented Gradient Maps based Representation ............................................. 141
  7.5.1 Representation of Complex Neuron Response ........................................... 141
  7.5.2 Oriented Gradient Maps by Response Vectors ........................................... 143
  7.5.3 The Properties of Distinctiveness and Invariance ..................................... 143
  7.5.4 Design of the OGM based Keypoint Detector ........................................... 144
7.6 Local Feature Matching ............................................................................. 145
  7.6.1 SIFT-feature based Matching ..................................................................... 145
  7.6.2 Score Level Fusion ................................................................................. 146
  7.6.3 Illustration of Matching Samples ............................................................. 147
7.7 Experiments ............................................................................................... 147
  7.7.1 Effectiveness of Multi-Level Keypoint Detection ....................................... 149
  7.7.2 Discriminative Power of OGMs .................................................................. 150
  7.7.3 Impact of Gallery Size ............................................................................. 151
  7.7.4 Verification Validation ............................................................................. 152
  7.7.5 Comparison with the State of the Art ....................................................... 155
  7.7.6 Complementarity of Left and Right Hands ............................................... 156
  7.7.7 Evaluation on Generalization Ability ........................................................ 156
  7.7.8 Complexity Analysis .............................................................................. 159
7.8 Summary and Discussion ............................................................................ 159

8 Conclusions and Future Work ....................................................................... 163
  8.1 Contributions .............................................................................................. 163
    8.1.1 3D face recognition based on principal curvature measures ...................... 163
    8.1.2 3D face recognition with heterogeneous cross-resolution data .................... 164
    8.1.3 Shape analysis based anti-spoofing face recognition ................................ 165
    8.1.4 Hand-dorsa vein recognition .................................................................... 165
  8.2 Perspectives of Future Work ....................................................................... 165
    8.2.1 Remeshing method in heterogeneous cross-resolution face recognition ........ 165
    8.2.2 Deep learning applied in 3D face recognition .......................................... 166

9 Publications .................................................................................................. 167

Bibliography ...................................................................................................... 169
# List of Tables

1.1 List of public 3D face database ............................. 18

4.1 Rank-one recognition rate achieved on FRGCv2 database for evaluating the impact of region size of Borel subset. ............... 75

4.2 Rank-one recognition rate achieved on FRGCv2 database for evaluating the impact of encoding pattern. ....................... 77

4.3 Rank-one identification rate of proposed LPCMP feature descriptor with three score-level fusion rules: the minimum fusion (the upper), the mean fusion (the middle), and the dynamic fusion (the lower). 78

4.4 Comparison of the state-of-the-art performing on FRGCv2 database. 81

5.1 The distribution of 3D face scans over the various probe subsets in the Bosphorus database. ................................. 96

5.2 The baseline identification rates of 3D FR with individual principal curvature measure based feature descriptor on FRGCv2 database . 97

5.3 Recognition performance evaluation of different fusion combinations of PCM-meshSIFT feature descriptor on the FRGCv2 database. Three score-level fusion rules are listed: Mean rule (upper sub-table), Product Rule (middle sub-table), and Minimum rule (lower sub-table). 99

5.4 Performance comparison on the FRGC v2.0 database. .......... 101

5.5 Rank-one recognition rate with pose and occlusion variations on Bosphorus database. ..................................... 102

5.6 Performance comparison on the subset of the Bosphorus dataset with various pose variations. 104

5.7 Performance comparison on the subset of the Bosphorus dataset with various occlusion. 104

5.8 Computation time consumed by each step in PCM-meshSIFT based 3D face recognition system .......................... 105

5.9 The rank-one recognition rates with the high-resolution samples in Lock3DFace database ............................ 114

5.10 The rank-one recognition rates with the low-resolution samples in Lock3DFace database .......................... 114

5.11 The rank-one recognition rates of heterogeneous cross-resolution based 3D FR on Lock3DFace database .................. 115

5.12 Computation time consumed in the online recognition. ........ 116

6.1 Configurations of the experimental scenarios .................. 124
### List of Tables

6.2 Verification and anti-spoofing performance evaluation ................................ 124  
6.3 Comparison of verification performance with spoofing attacks in Morpho database. ................................................................. 127  

7.1 Comparison of different local feature detectors. .............................. 138  
7.2 The results of different detectors, *i.e.* Harris-Laplace and Hessian-Laplace and their different fusion schemes for the multi-level keypoint detection based method on NCUT Part A. ........................................ 149  
7.3 Performance of each OGM and their combination in the setup of left-hand only, right-hand only and both-hands on the NCUT Part A database. ................................................................. 150  
7.4 The rank-one recognition rates of keypoint matching in each source as well as their combination, *i.e.* multiple sources, with respect to the gallery size of each subject on NCUT Part A. ........................................ 153  
7.5 Performance in the scenario of verification of the proposed method on the NCUT Part A dataset. ................................................................. 154  
7.6 Comparison with the state of the art in rank-one recognition rate on the NCUT Part A dataset. ................................................................. 155  
7.7 Comparative summary of related work on dorsal hand vein based identification and verification on different databases. .............................. 157  
7.8 The results of left hand only, right hand only, and their fusion using different numbers of gallery samples on the NCUT Part A dataset. ................................................................. 158  
7.9 The rank-one recognition rates of keypoint matching in each source as well as their combination, *i.e.* multiple sources, with respect to the gallery size of each subject on NCUT Part B. ................................................................. 158  
7.10 Performance in the scenario of verification of the proposed method on the NCUT Part B dataset. ................................................................. 158  
7.11 Average consumed time of each component of the dorsal hand vein recognition system. ................................................................. 159
List of Figures

1.1 Examples of human traits satisfying biometric selection criterion. . . . 3
1.2 Framework of a general biometric system. .............................. 3
1.3 Face recognition ability of human visual system performs regularly even though the resolution or the aspect ratio of face image is compressed. The individuals shown from left to right are: Michael Jordan, Zinedine Zidane, Barack Obama, Michael Jackson, Arnold Schwarzenegger, Woody Allen. .............................................. 7
1.4 Examples of face recognition related applications. ...................... 8
1.5 Schematic of generic face recognition system, which consists of four major modules: face detection, face alignment, feature extraction and face recognition. ..................................................... 8
1.6 Facial expression variations of Leonardo DiCaprio. ..................... 11
1.7 Comprehensive examples of head pose variations of Kobe Bryant. . 11
1.8 Illumination variations combined with head pose changes of the same subject [Sim et al. 2002]. ................................................................. 12
1.9 Examples of facial occlusion variations of Antonio Banderas. ....... 12
1.10 Examples of facial aging process of Audrey Hepburn. ................. 13
1.11 Examples of facial cosmetics. (a) and (b) respectively depict a subject without and with heavy makeup in YouTube Makeup (YMU) database. (c) depicts the subject without makeup, whereas, (d), (e) and (f) depict the makeup shots. They include the synthetic addition of (d) eye makeup, (e) lipstick and (f) full makeup. (c) (f) are generated by using Taaz software in Virtual Makeup (VMU) database. This figure is assembled from [Dantcheva et al. 2012]. .......................... 13
1.12 A scenario of structured light based 3D face scanning. A structured light pattern is shown in left, and the corresponding projection on face.[Tsakalakidou et al. 2005] ......................................................... 16
1.13 Examples of multi-view stereo based 3D face acquisition system proposed by [Beeler et al. 2010]. Left: Face model captured by a seven camera studio setup; Center: capture system; Right: Face model captured using consumer binocular stereo camera. ...................... 16
1.14 Illustration of the photometric stereo techniques. From left to right, raw image set captured under four direction lights; estimated field of surface normals in x, y, z channels; recovered depth map by performing integration on surface normals [Hansen et al. 2010]. ............. 17
1.15 Examples of different formats of 3D face scan, (a) depth image, (b) 3D shape model with texture, (c) 3D shape model in point clouds, (d) zooming eye corner region in point clouds, (e) 3D shape model in mesh, (f) zooming eye corner region in mesh. 19

1.16 3D face recognition processing flow. 20

1.17 Examples of high-resolution face scans and low-resolution face scans. A high-resolution meshed model contains 130,000 triangles, while a low-resolution meshed model contains 15,000 triangles. 22

1.18 Examples of the customers’ genuine face and the corresponding manufactured mask. (The figures are collected in “ThatsMyFace” website [ThatsMyFace]). 23

2.1 Example of ICP-based registration process including coarse alignment and fine alignment of 3D facial surfaces [Lu et al. 2006]. 33

2.2 The framework of the Thin-Plate Spline (TPS) based deformable model approach [Lu & Jain 2008]. 35

2.3 In left plot, expression normalization of two examples of the same subject. For the true face surface with noisy measurement in (a), the fitting algorithm gives a good estimation results in (b). The pose and expression normalized faces (c) are used for face recognition. In right plot, The reconstruction result shown in (b) is robust against the scans with artifacts, noise, and holes in (a) [Amberg et al. 2008]. 36

2.4 Three types of iso-level curves: iso-depth curves [Samir et al. 2006], circular curves [Samir et al. 2009], and radial curves [Drira et al. 2013]. 37

2.5 Extraction of a gallery region and three probe regions. (a) A gallery region, (b) probe region in the general central face area (probe C), (c) probe region in the nose region (probe N), (d) probe region in the interior nose region (probe I) [Chang et al. 2006]. 39

2.6 Left figure displays the facial muscle effect measurements on 3D shape of face, and the right figure shows the final split scheme and the segmentation scheme of 3D face [Amor et al. 2006]. 40

2.7 Left figure displays the 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. Right figure shows the cropping nose region in different sizes of radius [Faltemier et al. 2008]. 40

2.8 AvFM and its landmark points computed from FRGCv2 database (left image). Center and right images show seven facial regions and upper face region for the AvRM, respectively. The lower image displays separately the 7 meaningful regions [Alyuz et al. 2010]. 42
List of Figures

2.9 30 local regions and their corresponding rank-one recognition rate with individual regional classifier evaluated on FRGCv2 database [Spreeuwers 2011] ........................................... 43
2.10 Decomposition of face into sub-regions [Cook et al. 2006] .................. 45
2.11 The framework of Collective Shape Difference Classifier [Wang et al. 2010b] ........................................... 46
2.12 Illustration of remeshing procedure: original mesh with 13709 vertices, level-0, 2 and 5 remeshing samples with 7, 61 and 3169 vertices, respectively [Li et al. 2009] ................................. 48
2.13 Illustration of the generation of spin image and the related example [Johnson & Hebert 1999] .................................................. 49
2.14 Comparison of three 3D face representation [Mian et al. 2007] .......... 51
2.15 Ordered ring construction (first row) and encoding pattern of multi-resolution meshLBP (second row) [Werghi et al. 2016] ................. 51
3.1 The angle between two oriented triangles incident to \( e \) ..................... 63
4.1 The framework of 3D face recognition based on LPCMP-based feature descriptor .......................................................... 69
4.2 Illustration of the principal curvature measures computed on a 3D meshed face scan (A), and the corresponding curvature faces, the projected principal curvature measures in 2D image (B) .................. 70
4.3 Two examples of encoding pattern \( Q_{1,8} \) and \( Q_{2,16} \). The central gray point refers to the target point \( p_c \), and the black dots around refer to the neighbour points \( p_s \) .................................................. 71
4.4 Illustration of curvature faces (a ~ c) and the corresponding encoded curvature faces (e ~ f) with \( Q_{2,16} \) .................................................. 72
4.5 Examples of pre-processed 3D face scans in the FRGCv2 database. .... 74
4.6 Receiver operating characteristic (ROC) curves of True Accept Rate (TAR) to False Accept Rate (FAR) ............................................. 80
4.7 Detection Error Trade-off graph of False Accept Rate (FAR) to False Reject Rate (FRR) .................................................. 81
5.1 The framework of PCM-meshSIFT based 3D face description and recognition .......................................................... 87
5.2 The distribution of three principal curvature measures \( \lambda_{1B} \), \( \lambda_{2B} \) and \( \lambda_{3B} \) estimated on 3D meshed face. The second row displays the histograms of the value of principal curvature measures on the face. 88
5.3 Illustration of detected 3D keypoints using principal curvature measures \( \lambda_{1B} \), \( \lambda_{2B} \), and \( \lambda_{3B} \) (from left to right), respectively .......................... 89
5.4 The distribution of three principal curvature vectors $e_{1B}$, $e_{2B}$ and $e_{3B}$ estimated on 3D meshed face. ........................................ 91

5.5 3D keypoint descriptor configuration and representation. .................. 92

5.6 Samples of head pose variations in Bosphorus database. (a) frontal; (b) Yaw +10°; (c) Yaw +20°; (d) Yaw +30°; (e) Yaw +45°; (f) Yaw -45°; (g) Yaw +90°; (h) Yaw -90°; (i) Pitch upward rotation (PR); (j) Pitch downward rotation (PR); (k) Cross rotation (CR): yaw rotation and approximately pitch rotation +20°; (l) Cross rotation (CR): yaw rotation and approximately pitch rotation -20°. .......................... 94

5.7 Samples of external occlusion in Bosphorus database. (a) Occlusion by glasses; (b) Occlusion by hair; (c) Occlusion on eye part by hand; (d) Occlusion on mouth by hand. ...................................................... 95

5.8 Cumulative Match Characteristic of $HOC_1$, $HOC_2$, $HOC_3$ and their fusions from rank 1 to 10. ................................................................. 100

5.9 ROC curve with one individual principal curvature measures based feature and the fusion of three features. ................................. 100

5.10 Illustration of samples in Lock3DFace database. (a) neutral, (b) ~ (f) expression variations, (g) ~ (h) occlusion, (i) ~ (l) pose changes. ...... 107

5.11 Unweighted signed distance functions in 3D surface. (a) A sensor looking down the x-axis observes a range image, as a reconstructed range surface. Following one line of sight, a signed distance function is generated. The zero crossing point of this function is a point on the range surface. (b) When a new range surface incomes, the range sensor repeats the measures of the signed distance function for generating an optimized position of the point. The set of the points construct the iso-surface, as the optimized 3D surface. [Curless & Levoy 1996] 109

5.12 Illustration of reconstructed 3D face model with $n$ depth images (frames), $n$ is shown in the last row. ................................. 110

5.13 The keypoints distribute on the 2D face texture image. (a) Landmarks (blue points) located by Active Shape Model method; (b) Derivative keypoints (red points) located depending on the landmarks (blue points); (c) Eliminate the boundary landmarks (orange points) and combine the close points (green points). ................................. 112

6.1 2D texture image and corresponding 3D meshed scans in Morpho database. Samples in left column are genuine faces. In right column, the 1st and 3rd groups are type $A_A$ spoofing samples, while the 2nd group is type $A_B$ spoofing sample. ........................................ 123

6.2 Detection error tradeoff graph of experimental scenarios. .......................... 126
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Illustration of the NIR imaging system.</td>
<td>133</td>
</tr>
<tr>
<td>7.2</td>
<td>Hand-dorsa vein images: (a) from the NCUT Part A dataset and (b) from the NCUT Part B dataset.</td>
<td>133</td>
</tr>
<tr>
<td>7.3</td>
<td>Framework of the proposed approach, including comprehensive representation of optical properties through multi-order gradient quantities and robust matching with SIFT-based features.</td>
<td>136</td>
</tr>
<tr>
<td>7.4</td>
<td>The distribution of keypoints detected by (a) DoG, (b) Harris-Laplace; and (c) Hessian-Laplace (1,000 clusters), on a hand-dorsa surface. DoG locates very few feature points whose sums of $x$ and $y$ curvatures are extrema; Harris-Laplace identifies the keypoints whose elasticities are greater than a threshold; Hessian-Laplace detects the keypoints which carry shape information in terms of curvatures, and localizes in particular the ones which densely populate the valley regions corresponding to veins.</td>
<td>140</td>
</tr>
<tr>
<td>7.5</td>
<td>1000 selected keypoints located by the Hessian-Laplace detector: (a) based on clustering, and (b) based on strongest responses.</td>
<td>141</td>
</tr>
<tr>
<td>7.6</td>
<td>The neighborhood of the complex neutrons is a circular area and its radius can be changed according to the scale.</td>
<td>142</td>
</tr>
<tr>
<td>7.7</td>
<td>The OGMs describe a perceived Near-Infrared hand-dorsa vein image in 8 orientations.</td>
<td>143</td>
</tr>
<tr>
<td>7.8</td>
<td>Comparison in keypoint detection by DoG in the raw hand-dorsa vein image (center) and its corresponding OGMs in the eight pre-defined quantized orientations (around).</td>
<td>145</td>
</tr>
<tr>
<td>7.9</td>
<td>A matching example between the dorsal hand vein images belonging to the same person based on these keypoints detected using (a) Harris-Laplace and (b) Hessian-Laplace. The matched keypoints marked in yellow boxes are located in the vein region and the ones in red boxes are located in the nearby subcutaneous tissue.</td>
<td>148</td>
</tr>
<tr>
<td>7.10</td>
<td>A matching example using these OGM pairs of two left hands of the same person. The left column from top to bottom: OGM1 to OGM4; while the right column with the same order: OGM5 to OGM8. The matched keypoints marked in yellow are located in the vein area and the ones marked in red are detected in the nearby subcutaneous tissue.</td>
<td>148</td>
</tr>
<tr>
<td>7.11</td>
<td>Accuracy curves based on multi-level keypoint detection with respect to the gallery size of each subject on NCUT Part A.</td>
<td>152</td>
</tr>
<tr>
<td>7.12</td>
<td>Accuracy curves based on individual OGMs and their combination with respect to the gallery size of each subject on NCUT Part A.</td>
<td>152</td>
</tr>
<tr>
<td>7.13</td>
<td>Accuracy curves based on different sources of keypoint matching with respect to the gallery size of each subject on NCUT Part A.</td>
<td>153</td>
</tr>
</tbody>
</table>
List of Figures

7.14 CMC curves based on multi-source keypoint matching of different numbers of gallery samples of each subject on NCUT Part A. . . . . . 153
7.15 ROCs between FAR and VR of the proposed method on the NCUT Part A dataset. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154
7.16 ROCs between FAR and FRR of the proposed method on the NCUT Part A dataset. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154
7.17 Matching the keypoints across (a) scale variations and (b) translations (in both sub-figures, left column from top to bottom: Harris-Laplace, DoG on OGM$_1$, DoG on OGM$_3$, DoG on OGM$_5$, and DoG on OGM$_7$; right column from top to bottom: Hessian-Laplace, DoG on OGM$_2$, DoG on OGM$_4$, DoG on OGM$_6$, and DoG on OGM$_8$). . . 160
Face is one of the best biometrics for person recognition related application, because identifying a person by face is human instinctive habit, and facial data acquisition is natural, non-intrusive, and socially well accepted. In contrast to traditional appearance-based 2D face recognition, shape-based 3D face recognition is theoretically more stable and robust to illumination variance, small head pose changes, and facial cosmetics. The curvatures are the most important geometric attributes to describe the shape of a smooth surface. They are beneficial to facial shape characterization which makes it possible to decrease the impact of environmental variances. However, exiting curvature measurements are only defined on smooth surface. It is required to generalize such notions to discrete meshed surface, e.g., 3D face scans, and to evaluate their performance in 3D face recognition. Furthermore, even though a number of 3D FR algorithms with high accuracy are available, they all require high-resolution 3D scans whose acquisition cost is too expensive to prevent them to be implemented in real-life applications. A major question is thus how to leverage the existing 3D FR algorithms and low-resolution 3D face scans which are readily available using an increasing number of depth-consumer cameras, e.g., Kinect. The last but not least problem is the security threat from spoofing attacks on 3D face recognition system.

This thesis is dedicated to study the geometric attributes, principal curvature measures, suitable to triangle meshes, and the 3D face recognition schemes involving principal curvature measures. Meanwhile, based on these approaches, we propose a heterogeneous cross-resolution 3D FR scheme, evaluate the anti-spoofing performance of shape-analysis based 3D face recognition system, and design a supplementary hand-dorsa vein recognition system based on liveness detection with discriminative power.

In 3D shape-based face recognition, we introduce the generalization of the conventional point-wise principal curvatures and principal directions for fitting triangle mesh case, and present the concepts of principal curvature measures and principal curvature vectors. Based on these generalized curvatures, we design two 3D face descriptions and recognition frameworks. With the first feature description, named as Local Principal Curvature Measures Pattern descriptor (LPCMP), we generate three curvature faces corresponding to three principal curvature measures, and encode the curvature faces following Local Binary Pattern method. It can comprehensively describe the local shape information of 3D facial surface by concatenating
a set of histograms calculated from small patches in the encoded curvature faces. In the second registration-free feature description, named as **Principal Curvature Measures based meshSIFT descriptor** (PCM-meshSIFT), the principal curvature measures are firstly computed in the Gaussian scale space, and the extremum of Difference of Curvature (DoC) is defined as keypoints. Then we employ three principal curvature measures and their corresponding principal curvature vectors to build three rotation-invariant local 3D shape descriptors for each keypoint, and adopt the sparse representation-based classifier for keypoint matching. The comprehensive experimental results based on FRGCv2 database and Bosphorus database demonstrate that our proposed 3D face recognition scheme are effective for face recognition and robust to poses and occlusions variations. Besides, the combination of the complementary shape-based information described by three principal curvature measures significantly improves the recognition ability of system.

To deal with the problem towards heterogeneous cross-resolution 3D FR, we continue to adopt the PCM-meshSIFT based feature descriptor to perform the related 3D face recognition. Due to the adjustable region sizes of Borel subsets for estimating the principal curvature measures and vectors, the proposed descriptor is capable to match the face scans in different resolutions. We design a cross-resolution based recognition mechanism for practical system combining the discriminative power of high-resolution samples and the high efficiency of low-resolution samples. Specifically, the high-resolution 3D scans are registered as gallery set (i.e. offline part), while the low-resolution 3D scans perform the testing ones in probe set (i.e. online part). In order to optimize the low-quality facial surface, the Signed Distance Function based facial reconstruction algorithm is utilised to merger 60 frames of facial range image. Moreover, the keypoints consist to the landmarks located by a modified Active Shape Models based method and the supplementary keypoints determined according to these landmarks. Both the homogeneous and heterogeneous resolutions based experiments are performed on Lock3DFace database. The results demonstrate that our proposed method is able to achieve the promising matching rates in two scenarios effectively and efficiently.

In the research of anti-spoofing technology, due to the outstanding shape description ability of principal curvature measures, the difference between the surfaces of fraud mask and genuine face can be highlighted by PCM-meshSIFT feature descriptor. We adopt three categories of evaluation scenarios on Morpho database and FRGCv2 database, including baseline evaluation with all genuine faces in both databases; evaluation of anti-spoofing performance on Morpho database with high ratio of masks to genuine faces (10:9); evaluation of anti-spoofing performance simulating real-life testing case on both databases with low ratios of masks to genuine faces (5:95 and 1:99). A series of experimental results demonstrates that the shape-analysis based 3D face recognition approach is not only able to verify the person
identity with genuine face, but also to withstand the spoofing intrusion with mask manufactured.

We further study the hand-dorsa vein recognition. Due to the special data acquisition pattern that the hand-dorsa vein can only be imaged when the hot blood flows in vessels, it counts as an effective biometrics system to prevent the potential deceitful actions. Even though the vein vessels in back of hand are unique biometrics for each person, the sensors only provide low-contrast resulting image, which limits the distinctiveness in recognition. This thesis proposes a novel hand-dorsa vein recognition framework through matching local features of multiple sources. In contrast to current studies only concentrating on the hand vein networks, we also make use of person dependent optical characteristics of the skin and subcutaneous tissue, and encode geometrical attributes of their landscapes through different quantities. Specifically, the proposed method adopts an effective keypoint detection strategy to localize features on dorsal hand images, which is modeled as a combination of multiple order gradients. These features comprehensively describe the discriminative clues of each subjects. This method further robustly associates the corresponding keypoints between gallery and probe samples, and finally predicts the identity. Evaluated by extensive experiments, the proposed method achieves the best performance so far known on the NCUT Part A database. Additional results on NCUT Part B illustrate its generalization ability and robustness to low quality data.

**Key words:** 3D face recognition, Principal curvature measures, Heterogeneous cross-resolution, Anti-spoofing, Hand-dorsa vein recognition
Visage est l’une des meilleures biométries pour la reconnaissance de l’identité de personnes, car l’identification d’une personne par le visage est l’habitude instinctive humaine, et l’acquisition de données faciales est naturelle, non intrusive et bien acceptée par le public. Contrairement à la reconnaissance de visage par l’image 2D sur l’apparence, la reconnaissance de visage en 3D sur la forme est théoriquement plus stable et plus robuste à la variance d’éclairage, aux petits changements de pose de la tête et aux cosmétiques pour le visage. Spécifiquement, les courbures sont les plus importants attributs géométriques pour décrire la forme géométrique d’une surface. Elles sont bénéfiques à la caractérisation de la forme du visage qui permet de diminuer l’impact des variances environnementales. Cependant, les courbures traditionnelles ne sont définies que sur des surfaces lisses. Il est donc nécessaire de généraliser telles notions sur des surfaces discrètes, par exemple des visages 3D représenté par maillage triangulaire, et d’évaluer leurs performances en reconnaissance de visage 3D. En outre, même si un certain nombre d’algorithmes 3D FR avec une grande précision sont disponibles, le coût d’acquisition de telles données de haute résolution est difficilement acceptable pour les applications pratiques. Une question majeure est donc d’exploiter les algorithmes existants pour la reconnaissance de modèles à faible résolution collecté avec l’aide d’un nombre croissant de caméras consommateur de profondeur (Kinect). Le dernier problème, mais non le moindre, est la menace sur sécurité des systèmes de reconnaissance de visage 3D par les attaques de masque fabriqué.

Cette thèse est consacrée à l’étude des attributs géométriques, des mesures de courbure principale, adaptées aux maillages triangulaires, et des schémas de reconnaissance de visage 3D impliquant des telles mesures de courbure principale. En plus, nous proposons aussi un schéma de vérification sur la reconnaissance de visage 3D collecté en comparant des modèles de résolutions hétérogènes équipement aux deux résolutions, et nous évaluons la performance anti-spoofing du système de RF 3D. Finalement, nous proposons une biométrie système complémentaire de reconnaissance veineuse de main basé sur la détection de vivacité et évaluons sa performance.

Dans la reconnaissance de visage 3D par la forme géométrique, nous introduisons la généralisation des courbures principales conventionnelles et des directions principales aux cas des surfaces discrètes à maillage triangulaire, et présentons les concepts des mesures de courbure principale correspondants et des vecteurs de courbure
principale. Utilisant ces courbures généralisées, nous élaborons deux descriptions de visage 3D et deux schémas de reconnaissance correspondent. Avec le premier descripteur de caractéristiques, appelé Local Principal Curvature Measures Pattern (LPCMP), nous génerons trois images spéciales, appelée curvature faces, correspondant à trois mesures de courbure principale et encodons les curvature faces suivant la méthode de Local Binary Pattern. Il peut décrire la surface faciale de façon exhaustive par l'information de forme locale en concaténant un ensemble d'histogrammes calculés à partir de petits patches dans les visages de courbure.

Dans le deuxième système de reconnaissance de visage 3D sans enregistrement, appelée Principal Curvature Measures based meshSIFT descriptor (PCM-meshSIFT), les mesures de courbure principales sont d’abord calculées dans l’espace de l’échelle Gaussienne, et les extrèmes de la Différence de Courbure (DoC) sont définis comme les points de caractéristique. Ensuite, nous utilisons trois mesures de courbure principales et leurs vecteurs de courbure principaux correspondants pour construire trois descripteurs locaux pour chaque point caractéristique, qui sont invariants en rotation. Finalement, nous adoptons le classificateur à base de la représentation parimoniose pour l’appariement des points caractéristiques des deux visages. Les résultats expérimentaux basés sur deux bases de données, FRGC v2.0 et Bosphorus, démontrent que notre méthode proposée est efficace pour la reconnaissance de visage 3D et robuste contre des variations de pose et d’occlusion. De plus, la combinaison des trois mesures de courbure principale comportant l’information complémentaire améliore significativement la capacité de reconnaissance.

Pour résoudre le problème de reconnaissance de visages 3D en résolutions différentes et des sources hétérogènes, nous continuons d’adopter le descripteur caractéristique PCM-meshSIFT pour effectuer la reconnaissance de visage 3D associée. En raison des tailles ajustables des sous-ensembles Borel pour estimer les mesures de courbure principale, le descripteur proposé est capable de faire l’appariement entre des modèles en différentes résolutions. Nous proposons un système de reconnaissance basé sur un nouveau système pratique combinant la puissance discriminative des échantillons à haute résolution et l’efficacité élevée des échantillons à basse résolution. Plus précisément, les échantillons de visage 3D à haute résolution sont enregistrés d’abord (l’opération hors ligne), tandis que les échantillons de visage 3D à basse résolution effectuent le test (l’opération en ligne). Afin d’optimiser la surface faciale de qualité faible (basse résolution), l’algorithme de reconstruction de visage à l’aide de la méthode Signed Distance Function est utilisé pour fusionner 60 modèles de visage 3D dans la vidéo. De plus, les points caractéristiques sont constitués des points de repère déterminé par une méthode modifiée, Active Shape Model, et des points caractéristiques relatifs. Les expériences basées sur des échantillons homogènes et hétérogènes sont effectuées sur la base de données Lock3DFace. Les
résultats démontrent que la méthode proposée est efficace et efficiente pour vérifier des visages 3D aux résolutions distinctes.

Dans la recherche de la technologie anti-spoofing, en raison de la remarquable capacité de description proposée (PCM-meshSIFT), on est capable de distinguer des masques fabriqués et des visages authentiques par la différence mineure de formes de ses surfaces géométriques. Nous adoptons trois catégories de scénarios d'évaluation sur la base de données Morpho et la base de données FRGC v2.0, y compris l'évaluation de baseline avec tous les visages authentiques dans les deux bases de données; l'évaluation des performances anti-spoofing sur la base de données Morpho avec un ratio élevé de masques sur des visages authentiques (10: 9); et l'évaluation de la performance anti-spoofing simulant le cas de la vie réelle sur les deux bases de données avec de faibles rapports de masques à des visages authentiques (5:95 et 1:99). Les résultats expérimentaux démontrent que l’approche de reconnaissance de visage 3D basée sur l’analyse de la forme est non seulement capable de vérifier l’identité de la personne avec les visages authentiques, mais aussi de résister à l’intrusion avec masque fabriqué.

Enfin, Nous étudions la reconnaissance de vaisseaux veineux dans la main. En raison de la façon particulière d’acquisition de données que la veineuse ne peut être imagée que lorsque le sang chaud circule dans les vaisseaux, il est un système biométrique efficace pour prévenir les actions de tromperie potentielles. Néanmoins, même si les vaisseaux veineux à l’arrière de la main sont une indice biométrique unique pour chaque personne, les capteurs de réseaux veineux ne fournissent que des images résultant à faible contraste, ce qui limite les caractères distinctifs dans la reconnaissance. Cette thèse propose un nouveau cadre de reconnaissance des vaisseaux veineux dans la main en associant des caractéristiques locales de multiples sources. Contrairement aux études actuelles se concentrant uniquement sur les réseaux veineux de la main, nous utilisons également des caractéristiques optiques dépendantes de la peau et du tissu sous-cutané, et nous encodons les attributs géométriques de leurs paysages à travers différentes quantités. Plus précisément, la méthode proposée adopte une stratégie efficace de détection des points clés pour localiser les caractéristiques des images de la main dorsale, qui est modélisée comme une combinaison de gradients de plusieurs ordres. Ces caractéristiques décrivent en détail les indices discriminatifs de chaque sujet. Cette méthode associe plus étroitement les points de clés correspondants entre les échantillons pour prédire l’identité. Évalué par des expériences approfondies, la méthode proposée atteint les meilleures performances jusqu’ici connues sur la base de données de NCUT Partie A. Des résultats supplémentaires sur NCUT Partie B illustrent sa capacité de généralisation et sa robustesse aux données de faible qualité.

**Mots clés:** Reconnaissance de visage en 3D, Mesures des courbures principales, Résolutions Transversales Hétérogènes, Anti-spoofing, Reconnaissance de vaisseaux
veineux dans la main
1.1 Context and Motivation

1.1.1 Biometrics: An Innovation in Authentication
Authentication is an everlasting security topic focusing on a consistent effort to protect personal property, wealth, information and secret. A common authentication method is to verify a given artifact produced by a specific registered person as an authentication proof. In general, password, security question, person identification numbers (PIN), ID cards, driver license or personal tokens are popular authentication proofs. These authentication proofs have been widely used for a long time, because they are not expensive for production and take little trouble to be verified. However, the disadvantages of these proofs become apparent in cases...
of user’s memory loss, forgery or theft by impostors for a deceitful purpose. Therefore a more reliable and more secure authentication technology depending on the inherent factors of valid users was proposed and named as Biometrics. Biometrics is the science and the technology to measure and to analyze biological data of a person for authentication. Sometimes, it also means “Biometric Authentication” or “Biometric Recognition” in literature.

The biometric authentication proofs are commonly the biological, physiological, chemical or behavioural characteristics and traits related to a specific person. Seven main criteria to select biometric traits have been identified by Jain et al. [Jain et al. 2006] for the purpose of person authentication:

1. **Universality**: Each person possesses his/her own trait during an authentication process.

2. **Uniqueness**: The trait should be sufficiently unique for distinguishing one person from another.

3. **Permanence**: The trait should be invariant with one individual over a long period.

4. **Collectability**: A proper method or device can be applied to measure or capture the trait quantitatively.

5. **Acceptability**: The collection pattern or the measurement mode of the trait can be widely accepted by the public.

6. **Circumvention**: The vulnerability level of the underlying biometric system with the given trait is acceptable under fraudulent attacks.

7. **Performance**: The accuracy, processing speed and robustness of the biometric system involving the trait are sufficiently satisfactory to the authentication requests.

A number of biometric traits compelling more or less with the aforementioned criteria have been identified, including fingerprint, iris, face, hand veins, ear shape, palm print, voice, gait, retina, DNA, odor, signature, keystroke dynamics, etc. Examples of those traits are shown in Fig. 1.1. On the other hand, biometric traits, e.g., gender, age, height, weight, hair color or ethnicity, are shared by many different people and thereby cannot be used as authentication proofs because of their lack of uniqueness. Nevertheless they can help the person authentication process and have been labeled as “soft biometrics”.

In comparison with the traditional authentication proofs, these biometric traits show their advantages that they are obviously unforgettable and convenient to carry.
Figure 1.1: Examples of human traits satisfying biometric selection criterion.

Moreover, the biometrics system produces more barriers to impostors to forge the registered identity and relatively increases the security level. To study better the biometric system, some basic concepts will be presented as follows.

A standard biometric authentication process can be generally divided into three principal steps, including data capturing and preprocessing, feature extracting and template generating, and template comparing and matching (see Fig. 1.2).

Figure 1.2: Framework of a general biometric system.

Considering that one biometric system involving one qualified biometric trait, the first step is to capture information related to the trait from a group of valid
persons. For capturing samples of different biometric traits, the devices might be the digital camera, the NIR camera, the multi-spectral camera and the laser scanner, etc. Necessary preprocessing techniques are applied to improve the quality of captured data by removing noise, normalizing scale, aligning viewpoint, extracting region of interest (ROI), etc. Secondly, the system extracts features and generates templates from captured data. The templates are restored into disc as registered samples for later matching. This step is commonly called registration or enrollment. Finally, if a new sample (test sample) demands authentication, the system performs the first two steps and records the feature. The distance using a certain metric (e.g. Euclidean distance) between the feature of the registered sample and the one of the test sample should be measured. According to this distance, the matcher will output its authentication decision.

The function of biometric system can be roughly classified as “identification” and “verification” which are defined as follows.

- **Identification** It performs a one-to-all matching so as to determine one identity. The identity of the test subject is labelled as the one of the registered subjects if their distance is the minimum.

- **Verification** It performs a one-to-one matching so as to judge if the identity claimed is as same as the one of the test subject.

According to these two different scenarios, two series of metrics are used separately to evaluate the performance of a biometric system.

In the case of **identification**, the standard performance measurement of a biometric system uses the metrics listed as below.

- **Rank-1 Recognition Rate** It refers to the percentage of input template identified correctly. Only if the system labels the input template as the correct identity, the recognition process is successful.

- **Cumulative Match Characteristic (CMC)** It refers to a curve created by plotting the recognition rate against $n$-ranks. $n$-ranks indicates that the first $n$ enrolled templates most similar to input template are regarded as candidates. If the correct identity of the input template is in the candidate list, the identification result is correct.

In the case of **verification**, the standard performance measurement of a biometric system uses the metrics listed below.

- **True Accept Rate (TAR)** “True Accept” refers to the case that the biometric system correctly permits an access demand from a valid user with his/her own authentication proof. The rate means the percentage of valid inputs accepted correctly.
• **False Accept Rate (FAR)** “False Accept” refers to the case that the biometric system authorizes an access demand, but the authentication proof provided does not match to the identity claimed. The rate measures the percentage of invalid inputs incorrectly accepted.

• **False Reject Rate (FRR)** “False Reject” refers to the case that the biometric system refuses incorrectly an access demand from a valid person with right authentication trait. The rate means the percentage of valid inputs rejected incorrectly.

• **Receiver Operating Characteristic (ROC)** It refers to a curve created by plotting True Accept Rate against False Accept Rate at various threshold settings. It provides tools to select optimal verification models related to threshold settings.

• **Detection Error Tradeoff graph (DET)** It refers to a curve plotting False Reject Rate against False Accept Rate with a series of thresholds. It is a transformation function referring to ROC curve.

• **Equal Error Rate (EER)** A point in DET graph when False Reject Rate equals False Accept Rate. It provides a quick way to compare DET graphs generated by different verification systems. Generally, the system with a lower EER performs better.

The progress of the biometric technology is promoted and reinforced by the need for a large-scale identity management system with a higher accuracy guaranteed. Biometric applications have been spread in our modern society, which include sharing information via Internet, granting access to nuclear facilities, generating remote financial service, providing management of border entry and exit, etc [Jain et al. 2007]. Biometrics has reached a milestone in the authentication history.

### 1.1.2 Face: First Choice of Biometric Trait

With the long list of qualified biometric traits, a question raises naturally: which trait is the most suitable one to build up a biometric system? Throughout the human history, fingerprint, iris, signature, voice and face count as the top five most popular biometric traits. Apparently, identifying a person through a human face is the instinct of human beings. Then, In 1788 and 1949, people realized that fingerprint and iris are unique and distinctive to each person and used them to identify individuals respectively. Therefore, fingerprint, iris and face recognition technology gained most attention and developed rapidly in biometrics research area. While
recognition technologies based on fingerprint and iris have reached satisfying performance in commercial applications and products, their disadvantages are unveiled gradually along with their popularization.

- **Physical Intrusion:** Due to the fact that the relative position between the customer and the sensor is highly required for acquiring high-quality fingerprint or iris sample, customers need to pause for a second to declare themselves. This ‘pause and declare’ interaction is unfriendly and inefficient during high-throughput collection scenario [Pentland & Choudhury 2000]. Moreover, it hardly avoids a direct physical contact with a fingerprint or an iris acquisition device. This raises issues of how to keep the contact surface clean and germ-free during the acquisition process [Bowyer et al. 2006].

- **Social Intrusion:** Since people can’t recognize others by directly using fingerprint and iris data sample, these types of identification do not have a place in normal human interactions and social communication. This shortcoming is named as ‘oracle-like’ problem in [Pentland & Choudhury 2000]. It increases the difficulty to check the machine-based identification results by humans.

However, the face recognition technology sidesteps these “physical intrusion” and “social intrusion” problems naturally, because face data can be collected at a distance, and the subjects may not realize that they are scanned. The advantages of face recognition are concluded in the following points.

- **Intrusion-free** The face data acquisition process is unobtrusive (data acquired from a distance without ‘pause and declare’ interaction) and passive (various illumination conditions are accepted) to the test subject. During the process, there is neither waiting time, nor preset spot to stand by [Pentland & Choudhury 2000].

- **Natural and instinctive** Face recognition is the closest to humans’ perception habit and identification process. This biometric trait makes it easier for customers to accept and for a system manager to check recognition results.

In order to compare the performance of biometric identification involving different human traits in the case of flight boarding at airport, Hietmeyer obtained statistics of the compatibility of six biometric traits and reported them in [Heitmeyer 2000]. The statistics include face, fingerprint, hand geometry, voice, iris and signature based on the Machine Readable Travel Documents (MRTD) system. The compatibility score involves the factors of registration, renewal, machine-assisted identity verification requirements, redundancy, public perception, storage requirements and performance. As a result, the face recognition system achieved the highest score in compatibility which demonstrates that face is the first choice of biometric trait.
1.1.3 Overview of Face Recognition

Identifying other people by their face is an effortless task and an instinctive ability of human beings. Benefiting from outstanding adaptive ability of human visual system, human identification function performs well, even in some extreme situations. Humans are good at recognizing familiar faces even though there are natural changes in expressions, viewpoints and hairstyles [Pike et al. 2000]. Moreover humans can recognize faces in manipulated images, such as inverse images, low-resolution images and drastic compression images [Sinha et al. 2006] (as shown in Fig. 1.3).

![Figure 1.3: Face recognition ability of human visual system performs regularly even though the resolution or the aspect ratio of face image is compressed. The individuals shown from left to right are: Michael Jordan, Zinedine Zidane, Barack Obama, Michael Jackson, Arnold Schwarzenegger, Woody Allen.](image)

But the interesting phenomenon is that eyewitnesses to a crime can hardly describe and identify the perpetrator. The use of eyewitness evidence can also prove problematic as the witness may well express an inappropriate level of confidence in their decision [Pike et al. 2000]. A reasonable explanation is that crime victims were typically terrified by weapons holding by the perpetrator and ignored the detail information of the perpetrator’s face. However the machine-learning based face recognition technology is immunized to this emotional influence, and realizes an accurate suspect identity recording, tracking and recognising in and after the crime. Besides, thanks to the simplified operation process, the high social acceptance and the satisfying recognition accuracy of the face recognition technology, the related application has spread to numerous fields like photograph archiving, time atten-
dance, bank service, access control, boarder entry and exit, public safety, criminal investigation, etc.

Figure 1.4: Examples of face recognition related applications.

According to the conclusion in [Zhao et al. 2003], [Chellappa et al. 2010] and [Jain & Li 2011], a typical face recognition system consists of four major modules: face detection, face alignment, feature extraction and face recognition (see Fig. 1.5).

Figure 1.5: Schematic of generic face recognition system, which consists of four major modules: face detection, face alignment, feature extraction and face recognition.

Given an image or a recorded video captured by a camera, face detection module proceeds with detecting the face region and segments it from background. The face tracking technique is required to locate the face position in each frame of the video. Face alignment aims at normalizing scale, pose and/or illumination of face for simplifying subsequent module. The normalized face sample is transferred to feature extraction module, which transforms a face to certain measurable feature vectors with a discriminating representation. As the core module of a face recognition system, the selection of the facial feature is subject to several factors, including application purpose and environment, requirement of time cost and computation complexity, limitation of image quality and resolution, etc. Face recognition or face
matching is to match the features extracted from the testing face against those from enrolled faces in the database. The face recognition system finally outputs the identity of the testing face (identification) and/or the access authorization (verification) when a confident match is obtained.

1.1.4 Brief Review in 2D Face Recognition

Due to the simple data acquisition, the face recognition technology based on 2D facial image appears in ubiquitous applications as shown in Fig. 1.4. Ever since the first 2D face recognition work published by Kanade [Kanade 1973], researchers in the community of biometrics, pattern recognition and computer vision have been endeavouring to develop the 2D face recognition algorithms in recent decades. We introduce briefly the published algorithms in four categories: holistic feature based methods, local feature based methods, statistical based methods and deep learning based methods.

The design in holistic approaches derives from the face subspace theory, which assumes that the high-dimensional pixel arrays based representation of face images has intrinsic low-dimension structure. Both of Eigenface related Principal Component Analysis (PCA) [Turk & Pentland 1991a] and Fisherface related Linear Discriminant Analysis (LDA) [Belhumeur et al. 1997] are the representative approaches to extract the low-dimension feature in linear subspace from facial images. After that, a kernelization strategy was applied to transform the non-linear characteristics into linear ones by mapping the original face subspace to higher dimensional space using non-linear kernel function. Kernel based Principal Component Analysis (Kernel-PCA) [Yang 2002] and kernel based Fisher Discriminant Analysis (Kernel-Fisher) [Liu et al. 2002] are two typical methods. Besides, manifold learning strategy was exploited, which learns, identifies and parameterizes the intrinsic geometric structure of the non-linear face manifolds. The isometric mapping (ISO-MAP) [Tenenbaum et al. 2000] and Locality Preserving Projections (LPP) [He et al. 2005] are popular algorithms.

In contrast, local feature extraction methods analyse the texture information in partial region of the face, and represent compactly the feature of the whole face. The popular works are Gabor filter based [Wiskott et al. 1997], Scale-Invariant Feature Transform based (SIFT) [Bicego et al. 2006] and Local Binary Pattern based (LBP) [Ahonen et al. 2006] face recognition.

The third type of 2D face recognition approaches are based on the theory of statistical analysis of shapes. Active Appearance Model (AAM) [Edwards et al. 1998] aims to learn a model parameterized by several particular sets of attributes, and to seek the optimal set of model parameters which can generate a face image. These model parameters of each face image are regarded as feature to identify and verify
the faces. 3D Morphable Model (3DMM) [Blanz & Vetter 2003] is similar to AAM but the parameters are trained and generated from 3D laser scans of faces.

Face recognition via deep learning has achieved various breakthroughs in recent years. The Convolutional Neural Networks (CNNs) function as a classifier trains numerous parameters in different layers with abundant training data, and all the outputs from these layers can be used to construct the facial feature. Many researchers are interested in exploiting the face recognition methods involving the powerful CNNs [Parkhi et al. 2015] [Schroff et al. 2015] [Sun et al. 2013]. Chopra et al. proposed to train siamese networks for making the similarity metric to be small for positive pairs, and to be large for negative pairs [Chopra et al. 2005]. Taigman et al. proposed to supervise the learning process with CNNs by changing identification signal which brings richer identity-related information to deeply learned features [Taigman et al. 2014]. Joint identification-verification supervision signal is also adopted to gain more discriminative features [Sun et al. 2014][Wen et al. 2016].

1.1.5 Challenging Issues in 2D Face Recognition

In the past decades, the approaches to 2D face recognition have solved many authentication problems and achieved good recognition results. However, there still exist a series of challenging issues (e.g. expression, illumination, pose, cosmetics) related to 2D face recognition. These challenges appear principally in daily life where test samples are hardly controlled in an ideal condition. We summarize the challenging issues, e.g., facial expressions, head poses, illumination variations, occlusion, facial cosmetics as follows.

**Facial expressions** are facial changes in response to a person’s internal emotion states, intentions, or social communications [Jain & Li 2011]. It is the results of movement of the muscles beneath the face skin. Facial expressions observably change the geometry position of some local facial features and the topology structure of whole facial appearance (see Fig. 1.6). Eyes, eyebrows, lips and cheeks are typical non-rigid transformation regions on expressive faces, while forehead, chin and ears are rigid or semi-rigid regions. Except the neutral one, facial expressions are defined in six classes: happy, angry, fearful, surprise, sad and disgusted [Ekman 1993].

**Head pose** has three degrees of freedom: roll, yaw and pitch. Relative to the projection plane in camera, roll is an in-plane rotation treated as 2D problem which can be normalized and solved easily. But if there is an observable yaw and/or pitch rotation, deemed as out-of-plane rotation, some important facial features (e.g. eye corners or mouth corners) may be invisible in facial images. The combination of roll, yaw and pitch rotations of head immensely increases the difficulty in face recognition as shown in Fig. 1.7.

**Illumination variation** is another common environmental variable in 2D face
recognition. An intensive illumination variance may largely change the gray scale values in the facial image. A strong or weak light degrades the dissimilarity of gray scale in adjacent pixels and fades away the texture information. Furthermore, facial appearance is hard to distinguish if the change of head pose is fused into illumination variations (see Fig. 1.8).

Facial occlusions are often caused by a person’s hair, moustache, whiskers, hand,
Figure 1.8: Illumination variations combined with head pose changes of the same subject [Sim et al. 2002].

Figure 1.9: Examples of facial occlusion variations of Antonio Banderas.
Chapter 1. Introduction

sunglasses, sanitary mask, eyepatch, etc (see Fig. 1.9). These objects doubtlessly make parts of the face region invisible and result in degradation of face recognition performance.

Aging changes facial texture (e.g. wrinkles and age pigments) and facial shape (e.g. weight change and skull growth) [Park et al. 2010] [Ling et al. 2007] (as shown in 1.10). Facial aging is a complex problem in face recognition relating to passport and applications of finding missing children.

![Figure 1.10: Examples of facial aging process of Audrey Hepburn.](image)

Facial cosmetics also have clear impact on the performance of gray scale image based face recognition systems. Eye make-up is the most significant factor in the impact of facial cosmetics [Dantcheva et al. 2012]. Figure 1.11 gives two examples of facial cosmetics.

![Figure 1.11: Examples of facial cosmetics.](image)

Facial cosmetics also have clear impact on the performance of gray scale image based face recognition systems. Eye make-up is the most significant factor in the impact of facial cosmetics [Dantcheva et al. 2012]. Figure 1.11 gives two examples of facial cosmetics.

Apparently, all texture information and shape information in 2D facial images is provided by RGB and gray-level values, which are the major factors in 2D face recognition processing techniques. However the variation of facial expressions, head
poses, illumination and other challenging issues may observably change the information. If the distribution, gradient or visibility of pixel value alters without control, the robustness of facial features and the accuracy of matching results decrease simultaneously.

1.1.6 Opportunities for 3D Face Recognition

Facing such challenging issues in 2D face recognition technology, a natural idea springs up that to add one more dimension is to achieve much more opportunities. Hereby, several main opportunities using 3D face recognition technology are listed as follows.

1. **Popularization of 3D scanning techniques:** Since 3D technology offers a wide array of possibilities in almost every application region in entertainment and engineering, the invention and the innovation of 3D scanning techniques spread out rapidly in recent years. In the wave of 3D digital revolution, measurement, reconstruction, scanning and reproduction of 3D objects in real work become more convenient, more accurate, more efficient and more economical. Nowadays, many different types of 3D sensors are available to capture static and dynamic 3D data of a facial surface. 3D sensors are able to generate qualified 3D data with various resolutions for fitting different application purposes. In the next section “3D Face Acquisition Techniques”, more detailed introduction about 3D sensors will be presented.

2. **Potential solutions for challenges of 2D face recognition:** As states in [Medioni & Wapotitsch 2003], “Because we are working in 3D, we overcome limitations due to viewpoint and lighting variations”. Similarly, another paper states that facial data involving depth information “has the advantage of capturing shape variation irrespective of illumination variabilities” [Hesher et al. 2003]. The popular 3D sensors provide both texture and shape information of 3D face model. The supplementary shape information is less susceptible to variations in intensity image, because the geometric shape attributes on most parts of face are stable to illumination variance, head pose change, and facial cosmetics.

3. **More reliable features extracted depending on facial shape information:** In contrast to RGB or gray scale values in intensity images, the depth value is the precious information offered by 3D scan samples. The reasonable geometric attributes (e.g. Gaussian curvature, mean curvature, and normal) are available to produce more features which can precisely describe the shape of face. A paper states that “Depth and curvature features have several advantages over more traditional intensity-based features. Specifically, curvature descriptors: (1) have
the potential for higher accuracy in describing surface-based events, (2) are better suited to describe properties of the face in low contrast areas such as the cheeks, forehead, and chin, which are ignored in 2D face recognition, and (3) are viewpoint invariant” [Gordon 1992]. These facial shape based features guarantee the robustness to illumination and pose variations and the accuracy of recognition system in parallel.

1.2 3D Face Recognition

1.2.1 3D Face Acquisition Techniques and Databases

In recent decades, 3D scanning, capturing and reconstruction techniques have been largely developed and improved. In this section, three typical types of 3D face acquisition techniques are summarized as follows.

1. **Structured light:** Structured light is constructed by arrays of light with a predefined pattern of decided geometric relationship and shape. When the light is projected on facial surface, the transformation of the reflected light can be captured and measured to estimate the surface shape. Figure. 1.12 shows a predefined pattern of structured light. Therefore, the basic components of the structured light system include a light projector and a camera. Current structured light based 3D scanner can capture both static and sequences of 3D face scans in real-time. But, to guarantee a high acquisition quality, the position and the pose of face are restricted. The whole face should be covered by exposure range of structured light and perceive range of camera. Minolta Vivid 900/910 series, Inspeck Mega Capturor II 3D, and Kinect series are popular structured light based 3D scanner. Specially, Microsoft claimed that the depth measurement technique of Kinect depends on a random speckle light pattern instead of predefined light pattern [Garcia & Zalevsky 2008]. But here, they both count as the structured light based 3D face acquisition. Structured Light has been widely used in 3D face data acquisition in [Beumier & Acheroy 1999] [Huang et al. 2003] [Zhang & Huang 2004] [Zhang & Yau 2006] [Tsalakanidou et al. 2005].

2. **Multi-view stereo:** The multi-view stereo capturing technique utilizes several calibrated cameras positioning on spots with various viewpoints to face. One facial image is captured simultaneously by each camera, and the corresponding points in facial image from different cameras are used to estimate and reconstruct the shape of 3D face. The 3D reconstruction is operated off-line because of the huge computational complexity. During the acquisition process, the constant light without flash is sufficient, which is comfortable for person scanned. However, in [Bowyer et al. 2006], the authors state that accurate reconstruction of smooth
Chapter 1. Introduction

Figure 1.12: A scenario of structured light based 3D face scanning. A structured light pattern is shown in left, and the corresponding projection on face. [Tsakanikou et al. 2005]

surfaces (e.g. some face regions without enough nature texture) is very difficult to capture by using this technique. Latestly Beeler et al. published their enhanced multi-view stereo based acquisition system. The accuracy of reconstruction of facial surface has reached sub-millimetre level, which is comparable with the structured light based system [Beeler et al. 2010]. Multi-view stereo techniques are recently used in 3D face database building in [Yin et al. 2008], [Benedikt et al. 2010] and [Cosker et al. 2011]. The reconstructed 3D face scan and the capturing system based on multi-view stereo technique are displayed in Fig. 1.13.

Figure 1.13: Examples of multi-view stereo based 3D face acquisition system proposed by [Beeler et al. 2010]. Left: Face model captured by a seven camera studio setup; Center: capture system; Right: Face model captured using consumer binocular stereo camera.

3. Photometric Stereo: Photometric stereo is estimating the surface normal of objects by capturing a set of images of the object under different light conditions [Woodham 1980]. Fig. 1.14 displays an example of 3D facial surface reconstruction by four lights with different directions. Photometric stereo technique can capture 3D face sample in a high speed. The capture process is completed in 20 ms as reported in [Hansen et al. 2010]. But the quality of captured scan
Chapter 1. Introduction

is sensitive to the presence of projected shadows, highlights and non-uniform lighting. Furthermore, only the normal rather than the position of each point is estimated. It results in more computation complexity and more errors while reconstructing a facial surface. Photometric stereo system is used in constructing 3D face database introduced in [Zafeiriou et al. 2011].

Figure 1.14: Illustration of the photometric stereo techniques. From left to right, raw image set captured under four direction lights; estimated field of surface normals in $x$, $y$, $z$ channels; recovered depth map by performing integration on surface normals [Hansen et al. 2010].

Over the past decades, 3D face acquisition techniques have been used to construct various public 3D face databases, for example, FRGC v1.0, FRGC v2.0 [Phillips et al. 2005], BU-3DFE [Yin et al. 2006], Bosphorus [Alyüz et al. 2008], 3D-TEC [Vijayan et al. 2011], UMB-DB [Colombo et al. 2011b], GavabDB [Moreno & Sánchez 2004], SHREC’11 [Veltkamp et al. 2011], UND-DB [UND 2008], etc. In recent years, more databases are constructed by collecting 3D face data using Kinect: e.g. Florence-SuperfaceDB [Berretti et al. 2012], KinectFaceDB [Huynh et al. 2012], 3D Mask Attack database [Erdogmus & Marcel 2013], CurtinFaces [Huynh et al. 2012], FaceWarehouse [Cao et al. 2014], Lock3DFace [Zhang et al. 2016b]. Table 1.1 lists the devices, the number of subjects in total, the number of samples in total, and the variations of these databases. From this table, we observe that Kinect attracts more and more attentions in recent years. This capturing device costs less in purchase and utilisation, and thus makes it more suitable for some cases involving large scale of population.

1.2.2 Basic Concepts and Terminology

Face is a typical biometric trait, thereby the concepts and terminology in general biometric recognition system, as presented above, are also suitable to 3D face recognition research except for some specific cases. Therefore, a brief summary will be introduced in this section to clarify the basic concepts and terminology of 3D face recognition.

The general “3D face recognition” indicates two categories of application scenarios: one scenario is called “3D face identification”, and the other one is “3D face
Table 1.1: List of public 3D face database

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GavabDB</td>
<td>Minolta Vivid 700</td>
<td>61</td>
<td>549</td>
<td>P,E</td>
</tr>
<tr>
<td>FRGC v1.0</td>
<td>Minolta Vivid 910</td>
<td>275</td>
<td>943</td>
<td>-</td>
</tr>
<tr>
<td>FRGC v2.0</td>
<td>Minolta Vivid 910</td>
<td>466</td>
<td>4007</td>
<td>E</td>
</tr>
<tr>
<td>3D-TEC</td>
<td>Minolta Vivid 910</td>
<td>214</td>
<td>428</td>
<td>E,Twins</td>
</tr>
<tr>
<td>Bosphorus</td>
<td>Inspeck Mega Capturor II</td>
<td>105</td>
<td>4666</td>
<td>E,P,O</td>
</tr>
<tr>
<td>BU-3DFE</td>
<td>3DMD</td>
<td>100</td>
<td>400</td>
<td>E</td>
</tr>
<tr>
<td>BU-4DFE</td>
<td>3DMD</td>
<td>101</td>
<td>60600 v</td>
<td>E</td>
</tr>
<tr>
<td>BFM</td>
<td>ABW-3D</td>
<td>200</td>
<td>600</td>
<td>E</td>
</tr>
<tr>
<td>UND45-J2</td>
<td>Minolta Vivid 910</td>
<td>415</td>
<td>1800</td>
<td>P</td>
</tr>
<tr>
<td>UND-F</td>
<td>Minolta Vivd 910</td>
<td>302</td>
<td>942</td>
<td>P</td>
</tr>
<tr>
<td>UND-G</td>
<td>Minolta Vivd 910</td>
<td>235</td>
<td>738</td>
<td>P</td>
</tr>
<tr>
<td>3D-TEC</td>
<td>Minolta Vivid 910</td>
<td>214</td>
<td>428</td>
<td>E,Twins</td>
</tr>
<tr>
<td>SHREC’11</td>
<td>Roland/Escan scanner</td>
<td>130</td>
<td>780</td>
<td>P</td>
</tr>
<tr>
<td>UMB-DB</td>
<td>Minolta Vivid 900</td>
<td>143</td>
<td>1473</td>
<td>E,O</td>
</tr>
<tr>
<td>PhotoFace</td>
<td>photometric-stereo device</td>
<td>453</td>
<td>3187</td>
<td>Ti,E</td>
</tr>
<tr>
<td>FlorenceSuperfaceDB</td>
<td>Kinect</td>
<td>50</td>
<td>50 v</td>
<td>P</td>
</tr>
<tr>
<td>KinectFaceDB(Eurecom)</td>
<td>Kinect</td>
<td>52</td>
<td>936</td>
<td>E,I,P,O,Ti</td>
</tr>
<tr>
<td>3D Mask Attack DB</td>
<td>Kinect</td>
<td>17</td>
<td>255 v</td>
<td>Ti</td>
</tr>
<tr>
<td>CurtinFaces</td>
<td>Kinect</td>
<td>52</td>
<td>4784</td>
<td>E,I,P,O</td>
</tr>
<tr>
<td>FaceWareHouse</td>
<td>Kinect</td>
<td>150</td>
<td>3000</td>
<td>E</td>
</tr>
<tr>
<td>Lock3DFace</td>
<td>Kinect</td>
<td>509</td>
<td>5711</td>
<td>E,I,P,O,Ti</td>
</tr>
</tbody>
</table>

“Variations” labels: (P) Pose; (E) Expressions; (O) Occlusions; (I) Illumination; (Ti) Time. v denotes to the video clip samples.
verification”. As pointed out in [Bowyer et al. 2006], 3D face identification scenario is more challenging than 3D face verification scenario. Three principal reasons are listed as follows: (1) In identification scenario, a larger scale of gallery set might introduce more incorrect matching results; (2) For each test attempt in identification scenario, the whole gallery set should be searched in some way, which increases the time complexity apparently; (3) The verification scenario belongs to binomial classification, which is easier to guarantee a high recognition score; (4) In general verification scenario, subjects are assumed to cooperate with capture demands, while there are much more variations of head pose, illuminations, or occlusions in identification scenarios.

According to the settings of capturing device, both shape and texture information of 3D face scan is generally provided in the data acquisition process. A 3D face scan can be restored and displayed in several formats. In Fig. 1.15, four typical formats of 3D face scan are shown, and Fig. 1.15 (d) and (f) zoom in on the eye corner region to display the detail. Figure 1.15 (a) is a “depth image”, also sometimes called a “range image”, which provides the depth information of face in the 2D image. The pixel value reflects the distance from one point on face to the sensor that follows this rule: the bigger values are closer to the sensor, while the smaller values are farther away. When to display the 3D face scan combining texture and shape, the 2D image can be thought of a texture map overlaid on 3D shape model as shown in Fig. 1.15 (b). Figure 1.15 (c) shows the 3D shape model in point clouds, which records only the spatial coordinates of the scattering points on the facial surface. Figure 1.15 (e) shows the 3D shape model in mesh, which combines the spatial coordinates and the topology structure of all points on the face.

Figure 1.15: Examples of different formats of 3D face scan, (a) depth image, (b) 3D shape model with texture, (c) 3D shape model in point clouds, (d) zooming eye corner region in point clouds, (e) 3D shape model in mesh, (f) zooming eye corner region in mesh.

As illustrated in Fig. 1.16, a typical 3D face recognition system includes four principal steps: 3D face preprocessing, 3D face alignment, 3D face feature extrac-
Chapter 1. Introduction

The processing flow begins at the input face scans, including the ones in both gallery set and probe set. Even under ideal illumination condition, there are unexpected noise, holes and spikes on eyes and nose-wings regions, and hairy region (e.g. hair, moustache and beard) of scanned sample. “Holes” are essentially facial areas with missing data, and “spikes” are an outlier error in the data. Therefore 3D face preprocessing is required to fill the holes by interpolation method, and remove spikes by performing surface-based filters. After that, detecting and locating the landmarks (e.g. nose tip, eye corners, and mouth corners) are usually applied to obtain the face region. 3D face alignment normalizes the smooth face scan to a reference face orientation or in a common 3D coordinate system. Iterative Closest Point (ICP) [Zhang 1994] is the most popular algorithm for this step. Then, the shape-based features of normalized and aligned scan are calculated in 3D face feature extraction step. The features extracted from samples in both gallery set and probe set are transferred to “3D matching” model to compare their similarity. According to the function of system, the identification and/or the verification result is offered as the final output.

Figure 1.16: 3D face recognition processing flow.

1.2.3 Main Challenges

In the literature, 3D face recognition technology has been widely expected to provide solutions to overcome the challenges in 2D face recognition. Meanwhile, the accuracy of 3D face recognition keeps increasing with the appearance of more sophisticated algorithms and methods. However, when we review the basic mathematical theory applied in shape analysis related 3D face recognition, a potential theoretical problem appears in research which uses the face scan in discrete surfaces. Moreover, compared to the general high resolution face models, the 3D face scans with low resolution can be more beneficial to increase the processing speed of verification scenario, while the quality of this kind of data seriously affects the verification accuracy. It is required to find a verification mechanism which can exploit advantages of 3D face scans with a low resolution, and evade the shortcomings. Moreover, the
spoofing intrusion always threatens the security of 3D face recognition system. Especially, due to the development of 3D scanning and reconstructing techniques, it leaves more opportunities to imposters to easily attack the 3D face recognition system. It is also required to enhance and improve the security level and the stability degree of the anti-spoofing technology. Hereby, these four main challenges in 3D face recognition are presented in detail as follows.

1. **Shape analysis-based 3D face recognition in triangle mesh:** “Shape-based” information is the prominent factor of 3D face scan, which is subject to calculating many shape-based attributes on facial surfaces. Among the typical attributes of curved surface, surface curvatures are the most popular one which is used to represent and characterize the geometry structures. However, as presented above, 3D faces are usually stored into depth images, point clouds, or triangle meshes, which are discrete surfaces. Both point clouds and depth images are discontinuous, while triangle meshes are continuous but piece-wise smooth. In particular, the curvature estimation formulations for $C^2$ smooth surface in the classical differential geometry are not suitable for 3D faces. Therefore, finding a curvature estimation method based on triangle mesh and designing a suitable facial descriptor for recognising meshed 3D face scan are significant works in 3D face recognition research.

2. **Impact of pose and occlusion variations on 3D face recognition in the real world:** In the practical 3D face recognition application, the status of the testing individual is not always under-control. Especially in the identification case that the testing subject is uncooperative, criminal investigation, missing person retrieval, the variations of head pose and external occlusion have serious impacts on the performance of 3D face recognition. Specifically, when the testing person wears glasses or masks, or naturally puts hands on face, the external objects (i.e. glasses, mask, hand) make a part of face disappear unexpectedly and vanish some meaningful shape information. For head pose changes, three dimensional shape information is useful to normalize the direction of head, but the sophisticated normalization step consumes too much time. Therefore, it is expected to propose a shape-based descriptor effective for general face recognition and robust to pose changes and occlusions.

3. **3D face scans in high-resolution or low-resolution?** The “resolution” of 3D face scans refers to the number of vertices and facets contained in the meshed model. For the general case in face recognition, a 3D face meshed model with higher resolution provides more detailed shape information and higher recognition performance. But, as another side of a coin, the device for capturing the high-resolution face scan is expensive, and more triangles in one meshed model also increase the
computation complexity in data processing. It hinders the popularization of 3D face recognition technology in some overcrowding application case. To deal with this problem, a possible solution is to reduce the resolution of 3D face scan by using the samples captured by a consumer-level sensor (e.g. Kinect). However, the lack of detailed shape information in 3D face scans with low-resolution has serious impact on the quality of feature extracted and the accuracy of recognition results. This apparent contradiction about the selection of the suitable resolution is an unavoidable challenge in 3D face recognition. Fig. 1.17 displays two examples of high-resolution and low-resolution 3D face scans.

Figure 1.17: Examples of high-resolution face scans and low-resolution face scans. A high-resolution meshed model contains 130,000 triangles, while a low-resolution meshed model contains 15,000 triangle.

4. Spoofing attacks: “Spoofing attack” is defined as an intrusive act of deceiving a biometric system by presenting a fake evidence or copied biometric trait to obtain valid authentication [Nixon et al. 2008]. For attacking 2D face recognition system, imposters can show a photograph printed on paper or a video displayed on any tablet to deceive it. The fake photograph or video of the valid person might be captured by cameras, or simply collected via internet. Fortunately, the data acquisition mechanism of 3D face recognition system is able to stop effectively this type of spoofing attacks, because photographs and videos can not provide the depth information. However, 3D reconstructing and printing techniques mature rapidly in recent year, which bring more convenience to manufacture 3D objects, like “wearable mask”. Nowadays, a social public web-
site “ThatsMyFace” [ThatsMyFace] provides wearable 3D mask printing service with only one frontal photo (a side-view photo is an option). Figure 1.18 shows some examples of genuine faces and their corresponding masks manufactured from this website. According to the external form, even the human eyes hardly distinguish the high-quality mask from the genuine face. Therefore, the 3D mask attacks raise a new challenge to anti-spoofing technology in 3D face recognition.

Figure 1.18: Examples of the customers’ genuine face and the corresponding manufactured mask. (The figures are collected in “ThatsMyFace” website [ThatsMyFace])

1.3 Thesis Contributions

1.3.1 Methodologies

Shape analysis is an important method in 3D face recognition research, in which, choosing one or several representative geometrical attributes, and generating a discriminative feature based on these attributes, are two principal methodological steps. For the first one, because of the challenges described in the previous section, the traditional curvature estimation methods are unsuitable for triangle mesh based 3D face recognition. Fortunately, Morvan et al. generalized the point-wise curvatures for the discrete surface and proposed the concepts of curvature measures based on geometrical measure theory [Morvan 2008, Cohen-Steiner & Morvan 2003]. Referring to the theories associated to the generalized curvatures, this thesis adopts mainly “principal curvature measures” as the geometrical attributes to describe the shape of facial surface. The estimation method of principal curvature measures is based on the generalization of second fundamental form and the concepts of asymp-
Chapter 1. Introduction

totic cone proposed in [Sun & Morvan 2015]. This method calculates the curvature measures associated to a neighbour region of a point rather that the curvatures on the individual points on the original facial surface, or on the fitting surface. This method overcomes the limitation of the 3D meshed face for curvature estimation, and brings more opportunities to adopt and modify some existed 3D face features.

According to the characteristics of principal curvature measures presented above, in this thesis, we design two 3D facial feature descriptor based on the principal curvature measures and vectors. Both of them are effective in facial surface characterization and discriminative in face recognition. Besides, we also modify one of the feature descriptors, which is a registration-free one and is robust to pose and occlusion variations, for heterogeneous cross-resolution based 3D face recognition. Finally, this thesis evaluates the anti-spoofing performance of the proposed shape analysis based 3D face recognition in various scenarios. Moreover, hand-dorsa vein, as a living biometric trait in data acquisition, is a complementary system to heighten the anti-spoofing performance.

1.3.2 Main Contributions

In the following, the three main contributions of this thesis are summarized respectively, and they will be further presented in Chapter 3, 4 and 5.

a. Principal Curvature Measures related facial feature description to 3D Face Recognition

Curvatures of smooth surface are fundamental quantities in classical differential geometry. Similarly, curvatures of discrete surface (e.g. triangle mesh) play a fundamental role in 3D shape analysis and have been widely used to exploit the 3D face recognition algorithms. However, how to define properly and estimate accurately curvatures on discrete surface still perplex us. In this thesis, we study the estimation of curvatures on general triangle meshes and their relating application to 3D face recognition. In particular, by generalizing the conventional point-wise principal curvatures and principal directions to discrete surface, and we introduced the concepts of principal curvature measures and principal curvature vectors. The estimation process is coherent for discrete surface, because the second fundamental form $h$, and the principal curvatures $\lambda_1$ and $\lambda_2$ defined on smooth surface are replaced by their corresponding measure forms with strong mathematical theories. Particularly, the associated matrix $M$ to the second fundamental measures $\overline{h}$ has three eigenvalues (i.e. principal curvature measures) and three corresponding eigenvectors (i.e. principal curvature vectors) rather than two respectively. They are expected to offer more meaningful and complementary shape information. Based on these geometry quantities, we designed two different shape-based feature descriptors for 3D face recognition systems.
Chapter 1. Introduction

1. Principal curvature measures are combined with Local Binary Pattern (LBP) method to generate a new facial feature descriptor called **Local Principal Curvature Measures Pattern** feature descriptor (LPCMP). The maps of principal curvature measures are projected to generate three curvature faces. In one curvature face, the value of each pixel is one principal curvature measure of corresponding vertex. Then the feature descriptor based on the mini-patch-related histograms of curvature face is calculated, and used to match by Sparse Representation-based Classifier (SRC).

2. Principal curvature measures and principal curvature vectors are used to construct a SIFT-like 3D facial feature, named as **Principal Curvature Measures based meshSIFT feature descriptor** (PCM-meshSIFT). We build the Gaussian scale space on meshed face scan and compute the Difference of principal curvature measures to detect the keypoints with scale invariance. The keypoint based local feature descriptor guarantees the performance even if a part of face is missing (occlusions case). In keypoint description stage, the canonical direction based on principal curvature vectors is firstly assigned, which guarantees the rotation invariance of descriptor and the effectiveness when heads rotate. Secondly, Histograms Of principal Curvature measures (HOC) are calculated in a neighbour region of keypoints for representing the local facial features. Particularly, we generate three facial feature descriptor corresponding to three principal curvature measures. Finally, sparse representation based fine-grained matching strategy is utilized for 3D keypoint matching.

In the identification experiments with both kinds of features, the experimental results demonstrates that each of them carries on discriminative shape information. Furthermore, the fusion of the features generated individually by three principal curvature measures augments obviously the identification accuracy to 90.82% and 97.96% with two feature respectively on FRGCv2 database. The experimental results prove that our proposed schemes are competitive for 3D face recognition. Furthermore, with the PCM-meshSIFT feature descriptor, the experimental results achieved on the Bosphorus database prove that this method is effective for 3D face recognition with head pose changes and occlusions.

b. **Heterogeneous Cross-Resolution based Face Recognition** In considering the promising discriminative power of PCM-meshSIFT feature descriptor, we further evaluate its performance in the heterogeneous cross-resolution based face recognition. In the cross-resolution recognition scenario, the high-resolution 3D face model is registered as gallery set, and the low-resolution 3D face models perform as the probe samples in the on-line operation part. The purpose of this design is to put the time-consuming work with high-resolution samples in offline
stage, while the most efficient processing stage with low-resolution samples is performed in online stage.

The feature extraction and the matching scheme follow the 3D face recognition system involving PCM-meshSIFT feature descriptor. Particularly, to deal with the low-quality 3D face models captured by Kinect, we reconstruct the facial surface by merging 60 frames of face range image with the Signed Distance Function. And we replace the keypoints by the landmarks located by Active Shape Models in 2D texture image and the keypoints determined by these landmarks. The experiment part evaluated the performance of this 3D FR system in homogeneous and heterogeneous resolution based scenarios. In the homogeneous resolution based experiments, the proposed system achieves the recognition rates of 96.24% in the high-resolution based scenario, and 93.27% in the low resolution based scenario. Even though the disparity of the samples in different resolutions enlarge the intra-class distance during the heterogeneous cross-resolution based experiments, the recognition accuracy also reaches 83.37% by using the proposed 3D FR system.

c. Anti-spoofing Technology With the growth of face recognition, spoofing mask attacks attract more attention in biometrics research area. In recent years, the countermeasures based on texture and depth image against spoofing mask attacks have been reported, but the research based on 3D mesh sample hasn’t been studied yet. For filling this gap, in this thesis, we propose to exploit the anti-spoofing performance of the PCM-meshSIFT feature descriptor. Specifically, the estimated principal curvature measures and the related feature give us more opportunities to discover the shape dissimilarity between the genuine face and the fraud mask. In this work, the evaluation of anti-spoofing performance on the Morpho database and the FRGC v2.0 database. The previous one contains genuine face samples and fraud mask samples, and the later one only contains massive genuine face samples. The evaluation experiments are divided into three scenarios: 1. Baseline evaluation of verification with the Morpho and the FRGC v2.0 database respectively; 2. Basic anti-spoofing performance evaluation with only the Morpho; 3. To simulate the real-life testing case, the FRGC v2.0 database is enrolled as extension registration group to augment the ratio of genuine face samples to fraud mask samples. The evaluation results prove that our system can guarantee competitive verification accuracy for genuine faces and promising anti-spoofing performance against spoofing mask attacks in parallel.

d. Hand-dorsa vein recognition by matching local features of multi-source keypoints
Among the widely used biometric traits, the hand-dorsa vein is an emerging one for people identification. Particularly, the data acquisition of hand-dorsa vein is a liveness detection which can contribute a lot to raising the security level of the
biometrics recognition system. Because when infra-red light shines onto the back of the hand, the qualified hand-dorsa vein image can be only captured if the hot blood flows inside the vein vessels, which absorbs more infra-red radiation that the surrounding skin and subcutaneous tissue. In the captured hand-dorsa vein image, the vein vessel parts are darker, while the surrounding parts are lighter. However, the Near-infrared sensors only provide low-contrast resulting images, which do not offer sufficient distinctiveness in recognition particularly in the presence of large population. This thesis proposes a novel approach to hand-dorsa vein recognition through matching local features of multiple sources. In contrast to current studies only concentrating on the vein vessels, we also make full use of person dependent optical characteristics of the skin and the subcutaneous tissue revealed by NIR hand-dorsa images. We encode geometrical attributes of ridges and valleys through different quantities, such as cornerness and blobness, closely related to differential geometry. Specifically, the proposed method adopts an effective keypoint detection strategy to localize features on dorsal hand images, where the speciality of absorption and scattering of the entire dorsal hand is modelled as a combination of multiple (1st-, 2nd-, and 3rd-) order gradients. These features comprehensively describe the discriminative clues of each dorsal hand. This method further robustly associates the corresponding keypoints between gallery and probe samples, and finally predicts the identity. Evaluated by extensive experiments, the proposed method achieves the best performance so far known on the NCUT Part A dataset, showing its effectiveness. Additional results on NCUT Part B illustrate its generalization ability and robustness to low quality data.

1.4 Thesis Organization

The remainder of this thesis is organized as follows:

- In Chapter 2, we review the representative and successful approaches for 3D face recognition. They are categorized as the holistic feature based approaches, the local region-wise feature based approaches and the local point-wise feature based approaches.

- In Chapter 3, we introduce the concepts and the estimation method of principal curvature measures and principal curvature vectors.

- In Chapter 4, we introduce the Local Principal Curvature Measures Pattern (LPCMP) based 3D face recognition framework, and the related experiment performed on the FRGC v2.0 database.
Chapter 1. Introduction

- In Chapter 5, we present the Principal Curvature Measures based SIFT-like feature descriptor for 3D meshed face scan. The related system is tested on the FRGC v2.0 database and the Bosphorus database to evaluate its discriminative power in 3D face recognition and its robustness to pose variations and occlusion. Furthermore, the modified system performs the heterogeneous cross-resolution based 3D face recognition experiment on the Lock3DFace database.

- In Chapter 6, we study the anti-spoofing performance of the shape analysis based 3D FR approach against mask attacks. Comprehensive experimental scenarios are utilized to evaluate the performance in the extreme situation and in the simulating real-life application.

- In Chapter 7, we introduce the hand-dorsa vein recognition system, including data acquisition process, multi-source keypoint detector, SIFT-based keypoint descriptor, matching module, and comprehensive experiments on the NCUT hand-dorsa vein database.

- In Chapter 8, we conclude this thesis and propose the perspectives for future works.
Chapter 2

Literature Review: 3D Face Recognition

Contents

2.1 Holistic Feature based Approaches ................................. 29
  2.1.1 Subspace based approaches (PCA and LDA) ................. 30
  2.1.2 Iterative Closest Point (ICP) ................................. 32
  2.1.3 Deformable Model ........................................... 34
  2.1.4 Iso-level curve based approaches ............................ 36
2.2 Local Region-wise Feature based Approaches .................. 38
2.3 Local Point-wise Feature based Approaches .................... 43
  2.3.1 Geometry-texture descriptor based approaches ............. 44
  2.3.2 Geometry-shape descriptor based approaches ............... 47
2.4 Discussion ......................................................... 52
2.5 Summary ......................................................... 53

The emerging 3D scanning and capturing techniques are the technical impetus of 3D face recognition leading to numerous related approaches. Generally, 3D face recognition approaches can be classified, referring to the taxonomy of 2D face recognition, into two categories: i.e. holistic feature and local feature based approaches. In this chapter, we review the most significant and representative approaches in 3D face recognition following these two categories. Particularly, the massive local feature based approaches are presented in two sub-categories: i.e. the region-wise and point-wise approaches.

2.1 Holistic Feature based Approaches

Holistic feature based approaches work directly on the whole range image, triangle mesh or point-cloud of face scans. This kind of methods extract the facial feature or perform the similarity comparison associated to the whole face. The classical holistic feature based approaches presented in the

\subsection{Subspace based approaches (PCA and LDA)}

Referring to similar method in 2D face recognition, range image (depth image) is the principal data format of face sample in subspace based approaches. The difference between 2D and 3D FR is the value of pixel in image are RGB value and depth value. The range images can be represented as vectors, \textit{i.e.}, as points in a high dimensional vector space. Given a range image with a resolution $p \times q$, it can be mapped to a vector $x \in \mathbb{R}^{pq}$ by a lexicographic ordering of the pixel elements. Even though this high dimensional embeds, the natural constraints of the imaging process limit that the data will lie in a lower-dimensional manifold. The primary goal of the subspace based approaches is to represent and parameterize this manifold with some corresponding optimal criteria [Moghaddam 1999].

Let’s suppose a training set containing $N$ face samples, whose resolution is also $p \times q$. This training set can be also represented as a data matrix $X = (x_1, x_2, \cdots , x_i, \cdots x_N)$, where $x_i$ is the $i^{th}$ face vector of dimension $n = p \times q$. The mean vector of the training samples $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ is subtracted from each face vector for normalization.

Due to the property of gray image, the depth value of each pixel in range image is normalized in 256 gray levels. These pixels construct a face space and each range image represented as a vector in it. For analysing this face space, PCA and LDA can be applied in different representation pattern with basis vectors to represent high-dimensional space. If one projects the range images to the basis vectors, the corresponding coefficients account as the feature of the range image. The similarity measurement (\textit{e.g.} Euclidean distance or the cosine of the angle) between two face samples is computed directly in the feature space.

Generally, all the representations can be regarded as a linear transformation from the original range image vector to a projected feature vector defined in (2.1).

\begin{equation}
Y = W^T X
\end{equation}

where $Y$ is a $d \times N$ matrix of feature vector, $d$ is the dimension of the feature vector, and $W$ is the transformation matrix. Particularly, in this definition, $d \ll n$. 
Principal Component Analysis (PCA) aims to search the best $Y$ for the distribution of images, and to use $Y$ to define the subspace of range images [Turk & Pentland 1991b]. All face images in training group are projected into this subspace to find out a set of weights, which describes the contribution of each face vector in the face space. For identifying a testing face image, it is projected to the trained face space to achieve the corresponding set of weights. The comparison result of the weights of test sample and registered samples counts as the similarity measurements. This process of PCA is based on Karhunen-Loeve transformation [Kirby & Sirovich 1990]. If facial image elements are random variables, the image can be treated as a sample of a stochastic procedure. The basis vectors of PCA are defined as the eigenvectors of the $n \times n$ total scatter matrix $S_T$.

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \quad (2.2)$$

The transformation matrix $W_{PCA}$ is composed of the eigenvectors corresponding to the $d$ largest eigenvalues. Meanwhile, the eigenvectors are also known as eigenfaces. Commonly, the input images represented as a $n$-dimension vector is reduced to a feature vector of $d$ dimensional vector, because the other eigenvectors corresponding to smaller eigenvalues are generally noise.

Generally, PCA is an unsupervised approach, because the identity label information is not required in the construction of face space. In contrast to PCA, Linear Discriminant Analysis (LDA) aims to find an optimal way to represent the face vector space for maximizing the discrimination between different subjects. The identity label information is the essential part in LDA to improve the identification performance [Belhumeur et al. 1997].

The transformation matrix $W_{LDA}$ in LDA is defined as:

$$W_{LDA} = \arg \max_W W^T S_B W$$

where $S_B$ and $S_W$ are the inter-class and intra-class scatter matrix respectively, and formally defined as in (2.4) and (2.5):

$$S_B = \sum_{i=1}^{c} N_i(x_i - \mu)(x_i - \mu)^T \quad (2.4)$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (2.5)$$

where $N_i$ denotes the number of training samples in class $i$; $c$ is the total number of different classes; $\mu_i$ is the mean vector of the samples belonging to class $i$: $X_i$
denotes the set of samples in class $i$.

PCA and LDA are applied in the following publications related to 3D face recognition. Achermann et al. [Achermann et al. 1997] apply the PCA related eigenface approach in 3D face recognition. They report the 100% recognition rate on a small database including 24 subjects each of which has 10 images. Later, Hesher et al. [Hesher et al. 2003] explore PCA with different eigenfaces and image sizes. The testing database: FSU 3D face database contains 37 persons with 6 typical facial expressions. The experimental scenarios put multiple face images in gallery set, which hence gives more chances to samples in probe set to have a correct match. Heseltine et al. [Heseltine et al. 2004] evaluate the verification performance of eigenface-based approaches with various metric measurements. The verification accuracy is tested on a larger database with 330 facial images and reported as the Equal Error Rate (EER) which is 17.8%. According to the results, they claimed that the Mahalanobis distance is better than Euclidean and Cosine distance. Furthermore, they extend the scale of test set to 1470 range image of 230 subjects and examine the performance of LDA with same framework in [Heseltine et al. 2004] (i.e. multiple metric measurements used). The lowest EER with LDA is 15.3% by using Cosine distance. Al-Osaimi et al. [Al-Osaimi et al. 2009] propose to combine the Expression Deformation Model and PCA technique to perform the non-rigid 3D face recognition. The rank-1 recognition rate of 98.14% is reported on the whole FRGC v2.0 database under the neutral vs. all experimental settings. The verification rate is 98.35% and 97.73% at 0.01% FAR for scans under neutral and non-neutral expressions.

### 2.1.2 Iterative Closest Point (ICP)

The Iterative Closest Point [Besl & McKay 1992] algorithm attempts to align two 3D surface (i.e. a reference surface and a query surface) represented in point-cloud or mesh iteratively. To achieve this purpose, ICP finds firstly the closest point in the reference surface corresponding to each of the $n$ points in the query surface. Beginning with an initial estimate, the algorithm computes a sequence of rigid transformation $T_i$ until there is no additional improvement in mean square distance between the two surfaces. Implementing a k-d tree [Bentley 1975] data structure, the time cost of nearest neighbour searches is $O(\log(n))$. Specifically speaking, let us define the reference surface $M \triangleq \{m_i\}_{i=1}^{N_m}$ containing $N_m$ points and the query surface $P \triangleq \{p_i\}_{i=1}^{N_p}$ with $N_p$ points. The transformation of the query surface towards the reference surface is assumed to be a linear map which can be defined with a rotation matrix $R$ and a translation vector $t$. The optimal transformation parameters of ICP are the ones which can minimize the distance error between the points in transformed query surface and the corresponding closest points in
Chapter 2. Literature Review: 3D Face Recognition

Reference surface. The object function is given by (2.6).

$$\min_{R,t,j \in \{1,2,\ldots,N_m\}} \left( \sum_{i=1}^{N_T} ||Rp_i + t - m_j|| \right)$$ \hspace{1cm} (2.6)

The apparent drawback of ICP is the selection of the initial positions of the reference surface and the query surface. If they are not good enough, ICP probably finds a local minima as the improper solution. Thereby, the centroid of the two surfaces are commonly aligned before ICP to ensure an accurate matching result.

![Figure 2.1: Example of ICP-based registration process including coarse alignment and fine alignment of 3D facial surfaces [Lu et al. 2006].](image)

ICP is widely applied in registration and matching step in 3D face recognition. The ICP-based registration step is introduced as a key step to correct head pose as in [Kakadiaris et al. 2007][Li et al. 2009]. Figure 2.1 depicts a general ICP-based coarse to fine alignment process of 3D facial surface. Medioni and Waupotitsch [Medioni & Waupotitsch 2003] perform ICP to match facial surface in three dimensions. The experiments test the methods on the database contains 100 subjects, each of which possesses 7 scans with different poses. The EER is reported as “better than 2%”. Lu et al. [Lu et al. 2004] report the ICP-based approach on 3D face recognition, which assumes that the gallery face is more complete 3D facial surface and the probe face is a frontal view that is likely a subset of the gallery face. In the experiments with samples from 18 individuals including some variations of pose and expression, the recognition rate of 97% is achieved. Wang et al. [Wang et al. 2006b] propose to modify ICP as Partial ICP, which chooses a part of point pairs to calculate dissimilarity measure during registration of facial surfaces. It can reduce the negative impact of facial expression variation. The experiments perform on a database including 360 face models from 40 individuals to
demonstrate that the power of Partial ICP is better than original ICP algorithms.

### 2.1.3 Deformable Model

Due to the non-rigid parts of human faces, the facial expression variations can dramatically change the shape of facial surface. It will affect the performance of holistic feature based approaches mentioned above, *i.e.* PCA, LDA, ICP. Therefore, deformable model is proposed to describe directly the non-rigid transformation of face for filling in the shortcoming. Based on the training step by using representative facial samples, the deformable model can be used for two purposes. (1) Learning the deformation quantity between different facial expressions and thus indicating if it belongs to intra-class variations. (2) Learning the transformation between different facial expressions to enlarge the gallery set with more expressive face models or to convert an expressive probe face to a neutral one for recognition.

Lu and Jain [Lu & Jain 2005] propose Thin-Plate Spline (TPS) techniques to deal with the non-rigid transformation of facial surface, which is the extension work for rigid transformation with ICP in [Lu et al. 2004]. The proposed approach is appraised with a 100-subject database, including neutral samples registered as gallery set and the remaining neutral and smiling expressions samples perform as probe set. Particularly, the samples in gallery set are whole-head data, while the ones in the probe set are frontal view. Comparing the two works above, the errors between the smiling probe to neutral one after rigid transformation process are reduced apparently after the non-rigid deformation stage. For the 98 neutral and 98 smiling probes, the rank-one recognition rate arrives at 89%. Moreover, they proposed another 3D face recognition approach in [Lu & Jain 2006, Lu & Jain 2008] which is more stable and robust to facial expression changes. The approach includes four essential steps: (1) Constructing the 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. Fig.2.2 displays the entire framework. The core part, the control group, contains 10 subjects with seven typical expressions which are collected from the MSU database. The test samples are 877 facial scans of 100 subjects from FRGC v2.0 database. The experimental results prove the robustness ability of their proposed methods against facial expression changes.

Kakadiaris *et al.* [Kakadiaris et al. 2007] extend their previous work [Passalis *et al.* 2005] and propose an annotated deformable model approach for handling the facial expression in 3D face recognition. The pipeline of the method is formed as follows: (1) Building Annotated Face Model (AFM) by averaging a set of 3D facial meshes, and segmenting the face into anatomical areas by anthropometric landmarks; (2) AFM based face alignment includ-
Chapter 2. Literature Review: 3D Face Recognition

Figure 2.2: The framework of the Thin-Plate Spline (TPS) based deformable model approach [Lu & Jain 2008].

ing multiple stages from coarse to fine methods. The alignment steps contain spin image based alignment, ICP and Simulated Annealing on Z-Buffers in sequence; (3) AFM-based deformable fitting by Elastically adapted Deformable Model [Terzpoulos & Metaxas 1990, Metaxas & Kakadiaris 2002]; (4) AFM based geometry images extraction. (5) Wavelet-based geometry image analysis; (6) Face matching based on the Haar metric and the Pyramid metric. The recognition performance of this method is examined on FRGC v2.0 database. The rank-one recognition rate of 97.0% and the verification rate of 97.0% (with 0.1% FAR) are reported.

Amberg et al. [Amberg et al. 2008] introduce an expression-invariant 3D face recognition method by building and fitting a 3D Morphable Model (3DMM) [Blanz & Vetter 1999] separating the identity-related information and expression-related information. For each face scan with or without expression, it is iteratively fitted to the 3DMM by minimizing a cost function. According to the fitting results, the expressive face can be normalized to a neutral face and used for face recognition. Fig.2.3 shows some examples of neutralization results of expressive face scans. The similarity measurement between these neutralized faces is computed by the angle between face parameters in Mahalanobis space. Comparing to ICP based methods, the experiments carried out on the Gavab and UND datasets illustrate that the proposed recognition framework based on 3DMM improves the recognition accuracy.

Wei et al. [Wei et al. 2007] explore the 3D model sequences for identification in 3D face recognition. In order to deform the tracking model into a non-rigid facial surface, they propose to minimize the energy function based on the dissimilarity
Figure 2.3: In left plot, expression normalization of two examples of the same subject. For the true face surface with noisy measurement in (a), the fitting algorithm gives a good estimation results in (b). The pose and expression normalized faces (c) are used for face recognition. In right plot, The reconstruction result shown in (b) is robust against the scans with artifacts, noise, and holes in (a) [Amberg et al. 2008].

between the tracking model and the face scan. Finally, the 3D motion trajectories are calculated by vectors from the tracked points of the current frame to the corresponding points of the first frame with a neutral expression. The average value of the features from all deformed frames in one sequence is used as final representation for classification. This approach is examined on 600 sequences of 100 individuals, and the accuracy is 90.7%.

2.1.4 Iso-level curve based approaches

Iso-level curve based approaches extract a series of typical curves on facial surface and compute the similarity measurements of the shape of curves to estimate the identity of faces. These approaches utilized principally the shape analysis theory and Riemannian geometry theory based on the assumptions that the facial surface is a twice differentiable connected surface with zero genus. In the following part, three types of iso-level curves are introduced. Each of them follows the common processing pipeline which consists of spline fitting, arc-length parametrisation, and resampling. Finally, a set of uniformed and ordered facial curves represent the facial surface. The comparison of facial surfaces is converted to the one-to-one matching in curve-based level.

Samir et al. [Samir et al. 2006] design the iso-depth curves which are planar curves sampled from pose normalized 2.5D depth images. A series of slicing planes parallel to the xy-plane are determined at first. An iso-depth curve is extracted by concatenating the points situating in the same slicing plane. An example of
iso-depth curves is shown in Fig. 2.4(a). With multiple gallery faces, a nearest-neighbour classifier achieves accuracies of 92% and 90.4% on the FSU and UND dataset respectively.

Circular curves are proposed and applied in 3D face recognition in [Samir et al. 2009, Ballihi et al. 2012]. In contrast to iso-depth curves, the set of slicing planes is the intersection of the facial surface and a sphere centered at the nose tip with various radius $\lambda$. The circular curves are illustrated in Fig. 2.4 (b). The reported rank-one recognition rate in [Ballihi et al. 2012] is 85.65% on FRGC v2.0 database.

In contrast to the closed curves above, the radial curves utilized in [Ballihi et al. 2012, Drira et al. 2010, Drira et al. 2013] are a series of open curves starting from nose tip. The radial curves are the intersection lines of facial surface, and a group of slicing planes containing the nose tip and perpendicular to the $xy$-plane. Apparently, the normalization of face to frontal view is necessary to generate the radial curves. One example of radial curves is given in Fig. 2.4 (c). In [Drira et al. 2010], the radial curve based techniques achieves over 90% recognition rate on the whole Gavab database. In [Drira et al. 2013], the rank-one recognition rate of 97% is evaluated on FRGC v2.0 database. And in [Ballihi et al. 2012], the rank-one recognition rate on FRGC v2.0 database is reported as 89.04%, but it is improved to 98.02% by synthesizing the circular curves and the radial curves.

![Figure 2.4: Three types of iso-level curves: iso-depth curves [Samir et al. 2006], circular curves [Samir et al. 2009], and radial curves [Drira et al. 2013].](image)

The iso-level curve based approaches are interim methods between the holistic and the local feature based approaches. They can supply both curve-level and point-level one-to-one correspondence across facial surfaces. As a significant advantage of these methods, the generated natural metric is elastic for deformable and non-rigid facial surface. However, the drawbacks of the iso-level curve based approaches are the high dependence of the accuracy of normalisation and the location of nose tip. The missing of nose tip or nose region is a major problem for them. In fact, this case appears commonly with consumer-level capture device when the sample is too close to scan the nose region. A possible solution is to replace the nose region by the corresponding part in gallery set, but the alignment step to gallery set is obligatory.
Local region-oriented feature based approaches are proposed to handle local shape deformation caused by facial expression changes. The principal part of this kind of methods is to divide the facial surface into several sub-regions, and to perform the feature extraction and the feature matching on these separated sub-regions. Finally the similarity measurements calculated in these sub-regions are fused to be the matching scores. In the following section, five approaches based on different region segmentation methods and the associated 3D face recognition frameworks, including the publications [Chang et al. 2006, Chang et al. 2005, Amor et al. 2006, Faltemier et al. 2008, Alyuz et al. 2010, Spreeuwers 2011].

Chang et al. [Chang et al. 2006, Chang et al. 2005] propose an expression-robust face recognition method by matching multiple overlapping nose regions. The nose regions include a nose circle, a nose ellipse, and a region composed of just the nose (see Fig.2.5). Specifically, a raw 3D face scan with registered 2D face image is the system input. In the preprocessing step, a 2D skin detector is utilized to remove the non-skin regions and to extract the face region. Then, based on a quadratic surface fitting method for curvature estimation, they apply the curvature-based facial landmark detection to locate the nose tip (peak region), eye cavities (pit region), and nose bridge (saddle region). The gallery and probe local regions is extracted on these detected landmarks. As displayed in Fig.2.5, three different regions are highlighted around the nose tip as a probe (Fig.2.5(cd)), and a relatively larger regions is extracted as a gallery(Fig.2.5(a)). Thus, the ICP-based matching algorithm is necessary to compare each gallery-to-probe region pairs, and the Mean Square Error (MSE) measure is used to compute the corresponding similarity score. Moreover, it is proved in the experimental results that product and sum fusion rules guarantee the best performance. The experiments perform on the FRGC v2.0 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, the authors pointed out the remaining problem of handling expression variation, because there is about 10% degradation when a neutral face is matched to a face with varying expressions. From this paper, two important conclusions are obtained: 1. Using partial face (e.g. nose region) can result in a more accurate recognition even for the scenarios of neutral-to-neutral face matching. 2. Score fusion of multiple overlapping regions can massively improve the performance.

Ben Amor et al. [Amor et al. 2006] present an improved region-based approach for 3D face authentication and recognition. It consists of two stages: the offline
stage and the online stage. In the offline stage, they build a full 3D face database with neutral expression from three partial faces (frontal, left profile and right profile). Meanwhile a manual segmentation undergoes on the full face scan to divide the face surface in accordance to muscle deformation property. Specifically, they analyse the facial expression according to both anatomical theory and some empirical tests, and manually classify the facial muscles according to the degree of deformations they cause. Figure 2.6 shows the muscle effects measurements on face and the corresponding segmentation results. In the online stage, the probe facial scan is captured and only the face region is cropped out. Then, the probe model is compared to the full face database for compute the recognition scores. Particularly, they apply ICP to calculate the recognition score by a region-based similarity metric. The experiments perform on ECL-3D-face database including 50 full 3D faces as gallery and 400 2D test models as probe. The best reported rank-one recognition rate is 97.86%, and EER is 5.5% in verification scenario.

Inspired to these two valuable conclusions, Faltemier et al. propose a region set method for 3D face recognition [Faltemier et al. 2008]. This method divides the facial surface into 38 regions distributing over the whole face (as shown in Fig. 2.7). The centroids of these regions are defined by using $X$ and $Y$ offset values from the nose tip location, and with a single radius or multiple different radii. The
Figure 2.6: Left figure displays the facial muscle effect measurements on 3D shape of face, and the right figure shows the final split scheme and the segmentation scheme of 3D face [Amor et al. 2006].

Nose tip is automatically detected by combining two or three methods introduced in [Faltemier et al. 2006]. In the experiment section, they discuss the accuracy of the individual regions, the impact of the number of regions used for fusion, and the recognition result associated to the fusion rules. The results show that the scenario with 28 sub-regions achieves the best performance, and the Borda count and consensus voting methods yield better results than the standard sum, product and min fusion rules. This paper reports a rank-one recognition rate of 97.2% on FRGCv2 database in a standard experimental protocol.

Figure 2.7: Left figure displays the 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. Right figure shows the cropping nose region in different sizes of radius [Faltemier et al. 2008].
Another regional registration based algorithm for expression-robust 3D face recognition is proposed in [Alyuz et al. 2010]. As shown in Fig.2.8, this method consists of four essential parts: (1) Automatic landmark location; (2) ICP-based global and local rigid face registration; (3) Facial feature extraction; (4) Region classifier fusion. In the landmark location part, five fiducial points, including nose tip, left/right inner eye pits, and left-most/right-most points of the lower nose border region, are located using curvature-based methods. These points are then used for coarse face alignment by using the Procrustes Analysis. Their proposed local region based registration method, called as the Average Region Model (AvRM), is inspired by the Average Face Model (AvFM) [Gökberk et al. 2006, Gökberk et al. 2008]. The AvRM defines 15 local patches in 7 meaningful regions (nose, left/right eye, forehead, left/right check and mouth). Another overlapping region containing eyes, nose, and forehead are also defined as upper face regions. Two types of 3D shape quantities: the coordinates of facial surface points and the curvatures of facial surfaces, are considered as descriptor. The nearest neighbour algorithm is used to achieve identification result for face classification. Both score-level (sum and product rules) and abstract-level (plurality voting) fusion strategies are considered. FRGCv2 and partial Bosphorus databases with pose variations and occlusions are examined. For FRGCv2, the best rank-one recognition rate of 94.80% is achieved by fusing 16 individual classifiers (8 for each descriptor) with modified plurality voting fusion. For Bosphorus DB, the same fusion method also obtained the best rank-one recognition rate of 98.08% with fusing two descriptors and 8 local regions. To further improve the performance, statistical analysis (LDA) is performed on the coordinate based feature for dimensionality reduction, called as statistical coordinate feature. The LDA subspace for the test on FRGCv2 database is constructed by using FRGCv1 database, while a subset of Bosphorus database (643 scans of 20 subjects) is used to build LDA subspace for the test on the remaining part of Bosphorus (2265 scans of 85 subjects). Based on these experimental settings, the rank-one recognition rate of 97.51% and 99.31% are reported on FRGCv2 database and Bosphorus database respectively. However, the generalization ability of LDA-based approach is questioned, because the rank-one recognition rate on FRGCv2 database drops to 94.55% with the LDA subspace which is built by the subset of Bosphorus database. The significant advantage of this method is its processing speed. Compared to [Chang et al. 2005], only one single registration is sufficient for a probe face which saves more time because of AvFM-AvRM method. This method totally takes about 131 seconds for identifying one face from a gallery within 466 subjects in FRGCv2 database. Specifically, the time consuming includes: 11s for landmark localisation, 8s for curvature estimation, 10s for AvFM-based registration, 15s per region for AvRM-based registration, 3ms per regions for similarity computation.
Spreeuwers proposes a fast and accurate 3D face recognition methods which performs the face registration to an intrinsic coordinate system [Spreeuwers 2011]. The origin of this intrinsic coordinate system is the nose tip, and three planes of coordinate system are determined by the symmetry plane of face and the angle of the nose bridge. The RANSAC (RANdom SAmple Consensus) algorithms [Fischler & Bolles 1981] is utilized to perform the cylinder fitting for face region detection and the line fitting for the nose bridge estimation. The aligned point cloud is used to generate high-resolution range image by resampling, which is further normalized by spike removing, hole filling and face cropping. To deal with the facial expression variance, the author designs a series of 30 overlapping regions over the facial range image, which are shown as white areas in Fig.2.9. The vectorization form of each region constructs the feature vector, and their dimensionality if reduced by PCA followed by LDA. Then, a likelihood ratio classifier is used for regional feature classification. Finally, the majority voting based fusion strategy performs the decision-level fusion of all regional classifiers. The regional classifiers are trained with the Bosphorus database (only front-view scans without occlusion) and the 3DFace database, while FRGCv2 database performs as the testing one. Fig.2.9 gives the rank-one recognition scores for each regional classifiers trained by Bosphorus database. Based on these classifiers, the best rank-one recognition rate is reported as 95.9% with single classifier, and 97.9% with the fusion of 30 classifiers. The similar experimental results are achieved by using the classifiers trained by 3DFace database. Finally, a rank-one recognition accuracy of 99% is obtained by fusing the classifiers trained by both databases. The processing efficiency is out-
standing in this approach. It takes only 2.5s to identify a single probe sample in the gallery with 466 subjects from FRGCv2 database.

![3D Face Recognition](image)

Figure 2.9: 30 local regions and their corresponding rank-one recognition rate with individual regional classifier evaluated on FRGCv2 database [Spreeuwers 2011].

### 2.3 Local Point-wise Feature based Approaches

Local point-wise feature based approaches aim to extract the feature on the points on face rather than on the regions. This type of methods can be further classified into geometry-texture descriptor based approaches and geometry-shape descriptor based approaches. The obvious advantage of geometry-texture descriptor based approaches is to adopt the mature 2D image based methods to extract the feature descriptor. But the gray value based texture face image and the depth value based face image offer different information, which make the texture image and the range image have different characteristics and properties. Therefore, it is required to modify the 2D image based methods to fit such dissimilarity. These modified 2D images based methods will be presented as follows: Gabor filter [Cook et al. 2006, Wang et al. 2010b], Log-Gabor filter [Cook et al. 2006, Cook et al. 2007], Haar wavelet [Cook et al. 2007, Wang et al. 2010b], Discrete Cosine Transform (DCT) [Cook et al. 2007], Local Binary Pattern (LBP) [Wang et al. 2010b, Huang et al. 2012b] and Scale Invariant Feature Transform (SIFT) [Huang et al. 2012b]. The geometry-shape descriptor based approaches analyse the geometrical shape-oriented information in 3D face samples, and generate the shape-based feature for recognition. Most shape-based features are constructed by the geometric attributes of the sur-
face. In this section, we will shortly review the curvature-based descriptors: spin image [Kakadiaris et al. 2007, Conde et al. 2006, Mian et al. 2007], point signature [Chua et al. 2000], local shape map [Wu et al. 2004], 3D shape context [Frome et al. 2004, Berretti et al. 2011], spherical/tensor-based face representation [Mian et al. 2007, Mian et al. 2008], and mesh-LBP [Werghi et al. 2015, Werghi et al. 2016].

2.3.1 Geometry-texture descriptor based approaches

Cook et al. [Cook et al. 2006] propose to perform Log-Gabor filter on range image to achieve the local facial descriptor in 3D face recognition. Compare to traditional Gabor filter, Log-Gabor filter can capture more information in the high frequency areas and also has desirable high-frequency 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 respectively). Furthermore, the authors define 49 square regions (7 × 7 grid) with 50% overlap in both horizontal and vertical directions. These regions are further decomposed by the filters in 3 scales to generate 147 sub-regions. This process is displayed in Fig. 2.10. PCA is applied to the Log-Gabor filter in each of 147 sub-regions for dimensionality reduction, and Mahalanobis Cosine distance measure is used for similarity computation. Finally, the score level fusion strategy with sum rule combines the classification results. The method is evaluated on FRGCv2 database. The reported the rank-one recognition rates are 92.93% in the “1st vs. the remaining samples” experimental protocol, and 94.63% in the “1st neutral sample vs. the remaining samples” experimental protocol.

With the similar framework, Cook et al. extend their work in [Cook et al. 2007]. The new contribution in this paper is that three multi-scale feature descriptions (Haar wavelet transformation, Log-Gabor filter, and Discrete Cosine Transform, respectively) are applied in 3D face recognition. Then, PCA and LDA techniques are applied for dimensionality reduction. Similarly to the previous one, a standard nearest neighbour scheme with the Mahalanobis cosine metric is applied, and the un-weighted summation of distances crossing all frequency bands count as the final matching score. The authors give two conclusions in the paper: 1. A multi-scale partitioning of the frequency domain can improve the discrimination power of both the PCA and the LDA subspace projection techniques, and 2. the over-complete representation using Log-Gabor filters provides a better partitioning of the space/frequency domain of recognition than regular Gabor filters, DCT, or the wavelet transformation.

Wang et al. [Wang et al. 2010b] propose a 3D face recognition method, called Collective Shape Difference Classifier (CSDC). As shown in Fig. 2.11, the method
Chapter 2. Literature Review: 3D Face Recognition

Figure 2.10: Decomposition of face into sub-regions [Cook et al. 2006].

consists of offline training and online testing. Three main stages are contained in offline training phase: (1) Preprocessing and posture alignment; (2) Sign Shape Difference Maps (SSDMs) computation and description; (3) Boosting training of the Collective Shape Difference Classifier (CSDC). The well-trained strong classifiers are used to face verification and identification directly. Specifically, after the input of a raw face scan, three Gaussian filters are applied to remove spikes, fill holes and smooth the data. A face region is detected and cropped via texture channel, and the corresponding 3D points are labelled as face points. In the posture alignment, facial symmetry plane is determined by registering a pair of mirrored face scans, and the nose tip is precisely located on the facial central profile. According to the nose tip and the profile, nose bridge direction is further estimated. After these steps, a standard intrinsic standard coordinate system is defined by the nose tip, nose bridge direction, and the normal of the facial symmetry plane. All the points on facial surface are transformed to this standard coordinate system for fine posture alignment. The aligned face scan is projected to generating a normalized range image. For matching two range images $I_1$ and $I_2$, their SSDM $D_s$ is defined as $D_s(i,j) = I_1(i,j) - I_2(i,j)$. The SSDM records not only shape differences, but also shape change patterns. Three local feature are adopted to describe different traits of SSDM, including Haar-like features, Gabor features, and multi-block LBPs. Haar-like features encode the change pattern of the shape difference by measuring the difference between region averages of SSDM. Gabor feature encodes the characteristics of a set of spatial localities and orientations of SSDM. And LBP features highlight the texture characteristics of SSDM. The proposed classifier
CSDC is defined as \( H_T(D_s) = \sum_{t=1}^{T} c_t(D_s) \), where \( c_t(D_s) \) is a weak classifier based on the local features on the boosting training of SSDMs. To training the CSDC, inter-class and intra-class SSDMs are built by given 3D face models from Bu-3DFE database. Three CSDCs are achieved: \( H_T^{Haar} \), \( H_T^{Gabor} \) and \( H_T^{LBP} \) refer to three local features. The comprehensive experiments evaluate the performance of the methods on FRGCv2 database. With FAR of 0.1%, the paper reports that the verification rates are 97.97%, 98% and 98% for ROC I, ROC II and ROC III masks respectively. However, only \( H_T^{Haar} \) is used in the face identification protocol. Rank-one recognition rates are reported as 98.3% and 98.4% respectively for the first sample vs. the remaining scenario and the first neutral vs. the remaining scenario. The paper also reports that to recognize one face scan against 1,000 faces takes only 3.6 seconds.

For extending the previous works presented in [Huang et al. 2011b, Huang et al. 2010], Huang et al. propose an extended Local Binary Pattern (eLBP) based facial description and local feature hybrid matching for 3D face recognition [Huang et al. 2012b]. To deal with a raw face scan, the basic preprocessing, including spike removal, smoothing and hole filling, is the first stage to optimize the input sample. Then, In contrast to the original LBP, the proposed eLBP contains two parts: 1) computing the relative gray value difference between central pixel and its neighbours, as LBP feature; 2) calculating the absolute differences between the central pixel and the surroundings, which are also important to describing the detailed information of local shapes. Similarly to LBP, 8 scales of eLBP are encoded, but 4 layers involving absolute differences are computed.
Chapter 2. Literature Review: 3D Face Recognition

Each original depth image can be finally represented by 32 multi-scale processed eLBP images. Then SIFT features [Lowe 2004] are extracted from these eLBP images for keypoint detection and feature configuration. The paper also proposed a hybrid matching process, which combines a local matching step with SIFT-based features and a global matching step under the facial component and configuration constraints. Finally, three matching similarity measurements are considered: the quantity of matched keypoint pairs, the similarity of the facial component constraint, and the similarity of the facial configuration constraint. Three public databases are tested, i.e. FRGCv2 database, Bosphorus database, and Gavab DB. A rank-one recognition score of 97.6% is reported on FRGCv2 database, a rank-one recognition score of 97.0% is achieved on a subset of Bosphorus database (3301 scans of 105 subjects with different expressions and occlusions), and the rank-one score is 91.39% for Gavab DB. All the experimental results demonstrate that the proposed eLBP feature is robustness to facial expression variations, external occlusions, and extreme pose changes.

2.3.2 Geometry-shape descriptor based approaches

Gordon counts as the first author to use curvature information to handle 3D face recognition in [Gordon 1992]. The author firstly proposed some designs and ideas of the use of curvature information. 1) Sign of Gaussian and mean curvature can be used to segment face into convex, concave and two types of saddle regions; 2) extrema values of two principle curvatures are used to extract rigid and valley lines on face surface; 3) principal directions can detect umbilic facial points; 4) curvature maps are able to detect facial features, e.g. nose region, nose bridge, eye corners, eyeballs. The detected facial features are beneficial to face pose normalization. The experiments are carried out on a test group of three views of each of 8 faces and recognition rates as high as 100% is reported. Similarly to the previous idea, Tanaka proposed Extended Gaussian Image (EGI) to highlight facial rigid and valley lines for face representation and Fisher’s Spherical Correlation to be similarity measurement in [Tanaka et al. 1998]. 37 facial range images are used to test the identification performance by the shape information from surface curvatures.

For establishing one-to-one point correspondence between facial surfaces, a series of methods is proposed to fit the samples to a generic face model and to perform surface re-meshing or re-sampling. In [Alyuz et al. 2010], the face registration step adopts a generic face model for fitting all input face samples. Li et al. put forward a technique in 3D face recognition via surface re-meshing in [Li & Zhang 2007, Li et al. 2009]. They explore the use of multiple intrinsic geometric attributes, including angles, area, geodesic distances, curvatures. To obtain the one-to-one correspondence and compare these low-level features, each face scan is represented.
by a triangular mesh, in which 43 vertices in mask are detected manually on original face, and a uniform re-meshing procedure is implemented as shown in Fig. 2.12. For handling expression variations, they train different weights for each individual attribute and perform feature ranking. Based on a nearest neighbor classifier and a Sparse Representation Classifier (SRC), they display the effectiveness of their methods on the Gavab database and the FRGC database. In [Li & Zhang 2007], the recognition rate of 94.17% is reported in normal-reference (NR) scheme, and the 97.00% is achieved in Leave-One-Out (LOO) scheme. In [Li et al. 2009], they achieve the one-best-matching results in different protocols as follows: 96.67% for neutral sample, 93.33% for expressive ones, and 94.68% for overall samples.

Moreover, some local shape descriptors are designed to describe the local shape of facial surface by characterizing some basic geometric attributes (e.g. curvature, distance, area, angle) in a regular or parametric space. These compact feature vectors has also been commonly explored in 3D face recognition.

Spin image is proposed in [Johnson & Hebert 1999] and becomes a well-known 3D shape descriptor. Given an oriented point consisting of a point on surface and its normal direction, one defines a grid image with $n$ rows and $m$ columns rotating around the normal direction. All the points will be “captured” by the grid image and transferred to projected points in grid (see Fig. 2.13). If the projected points does not situate on the grid, the authors propose to adopt bilinear interpolation method to transfer the points. After the rotation and the interpolation, the grid image with project points is called a spin image. It utilizes the relative position of surrounding points to describe feature of the basic point. This feature has been adopted in face surface registration [Kakadiaris et al. 2007], facial interest point detection [Conde et al. 2006], and face recognition [Mian et al. 2007].

Point Signature [Chua & Jarvis 1997] is proposed to describe the local shape by a 3D contour curve of a basic point. Given a point $p$, one places a sphere of radius $r$
centred at \( p \). The intersection of the sphere with the surface is a 3D curve \( C \), whose orientation can be defined by an orthonormal frame formed by a normal vector \( n_1 \), a “reference” vector \( n_2 \), and their vector cross-product. Here, \( n_1 \) is the unit normal vector of a fitting plane through \( C \) to surface. Another plane \( P' \) is defined by translating the fitting plane to \( p \) in a direction parallel to \( n_1 \). The perpendicular projection of \( C \) to \( P' \) forms a new planar curve \( C' \) with the projection distance of points on \( C' \), which generates a signed distance profile. \( n_2 \) is defined as the unit vector from \( p \) to the projected point on \( C' \), who gives the largest positive distance. Each point \( p \) on \( C \) is characterized by: 1) the signed distance from itself to the corresponding point on \( C' \); 2) a clock-wise rotation angle \( \theta \) around \( n_1 \) from the reference direction \( n_2 \). In [Chua et al. 2000], they propose to use point signature for 3D face registration and recognition. Experiments are performed on 6 subjects with 4 different facial expressions. The results demonstrated the effectiveness of their method.

Local shape map is motivated by the previous spin image method and point signature [Wu et al. 2004]. It is a 2D histogram constructed by mapping 3D co-ordinates of surface’s points with a sphere centralized at this point into 2D space. The 2D histogram is quantized by two factors: 1) the distance from a basic point to its neighbour point; 2) the displacement of the neighbour point to the tangent plane of the basic point. This method is evaluated on SAMPL database consisting of 31 range images from 6 subjects, and the lowest EER for different parameters is 2.98% with AND-rule. The authors claimed that local shape map outperforms spin images for 3D face recognition.

3D shape context [Frome et al. 2004] is a straightforward extension of 2D shape context [Belongie et al. 2002] to 3D case. It describes the point distribution of a
local shape of surface which quantizes the shape information by a 3D spherical sampling space. Berretti et al. propose to use 3D shape context to generate the SIFT-like framework with HOG, SHOT and GH descriptors for 3D face recognition via coarse keypoint matching in [Berretti et al. 2011]. The experiments are carried out on BU-3DFE database and Florence face dataset. The comprehensive experimental results demonstrate that their proposed methods can handle the 3D face recognition with facial expression, large pose changes, and partial face matching.

**Spherical Face Representation (SFR)** is proposed by Mian et al. [Mian et al. 2007]. It is an efficient and effective multi-modal 2D/3D hybrid face recognition method. Specifically, SIFT-based descriptor is used for 2D face matching, while SFR descriptor and region-ICP (only eye-forehead region and nose region) perform the 3D face matching. SFR can be considered as the characterization of the point cloud of a face following a spherical set of bins centered at the nose tip (see Fig. 2.14(b)). To generate an $n$ bin SFR feature, the distance from all points to the nose tip is computed firstly, and then they are quantized into a 1-D histogram of $n + 1$ bins. The outermost bin is discarded since it is prone to errors. When examined on FRGCv2 database, the rank-one recognition of 96.2% is reported by using region-ICP algorithm. This accuracy is further improved to 97.37% by fusing the 2D-SIFT matching, 3D-SFR matching and the region-ICP matching. Mian et al. also proposed a similar **Tensor-based Face Representation** in [Mian et al. 2006], which characterizes the facial surface area into a predefined 3D grid space. Fig. 2.14(c) shows a $10 \times 10 \times 10$ tensor over a facial point cloud. In [Mian et al. 2008], they propose another multi-modal 2D/3D hybrid face recognition framework consisting of local keypoint detection, description and matching. 2D face recognition part is performed by 2D SIFT-based face recognition pipeline on texture images. For 3D face recognition part, the modified **tensor-based face representation** is adopted for local shape description and the descriptor is further projected to a PCA subspace for dimensionality reduction. Finally, graph based matching strategy is explored for local feature matching. The proposed method achieves the rank-one recognition rate of 96.1% on the whole FRGCv2 database.

**meshLBP** is proposed by Werghi et al. [Werghi et al. 2015] to encode the Local Binary Pattern (LBP) directly on a triangle mesh, and extract the local feature considering geometrical (e.g. mean curvature, Gaussian curvature, curvedness, and shape index) and photometric (e.g. gray level) information for face recognition [Werghi et al. 2016]. LBP is a well-know local feature descriptor in 2D image processing [Ojala et al. 1996, Ojala et al. 2002], which can describe the photometric feature in 2D image. The authors extend LBP to triangle mesh case via designing an effective encoding pattern, called **ordered ring**. As shown in Fig. 2.15, the construction of ordered ring begins at forming three ordered facet $f_{out1}$, $f_{out2}$ and $f_{out3}$, which are adjacent to the central facet $f_c$ and point outward. Then a se-
Chapter 2. Literature Review: 3D Face Recognition

Figure 2.14: Comparison of three 3D face representation [Mian et al. 2007].

Figures of sequence of $F_{gap}$ facets locating between each pair $\langle f_{out1}, f_{out2} \rangle$, $\langle f_{out2}, f_{out3} \rangle$ and $\langle f_{out3}, f_{out1} \rangle$ are addressed. The $F_{gap}$ facets have exactly one vertex on the initial 3-edge contour of the central facet $f_c$, and they are named so because they look like filling the gap between the $F_{out}$ facets. This step produces a ring of facets distributing in a circular pattern around the central facet $f_c$. By iterating this step with the outer edges of $F_{out}$, the new set of $F_{out}$ and $F_{gap}$ in next ring are built.

The experiments perform on BU-3DFE database and Bosphorus database, including expression, pose, and occlusion invariance. The comprehensive experimental results demonstrate that the meshLBP method fusing photometric and geometric information has discriminative power in 3D face recognition.

Figure 2.15: Ordered ring construction (first row) and encoding pattern of multi-resolution meshLBP (second row) [Werghi et al. 2016].
2.4 Discussion

After the development of 3D face recognition in the last two decades, the objective of 3D face recognition is to meet the requirements of the real-world applications, i.e., to provide higher recognition accuracy (with one registered face model in gallery set), to handle more environmental variances (facial expression, head pose rotation), to deal with partial face data (external occlusion, partial face missing in data acquisition, head pose change in large intensity), and to identify persons more efficiently.

For satisfying these requirements, the representative features are designed, examined and published in massive publications as presented in this section. Based on the taxonomy of 3D face recognition, local feature based matching generally outperforms holistic feature based matching, especially to deal with the facial expression variations in 3D face recognition. Nowadays, many milestone holistic matching methods are regarded as benchmarks for performance comparison or some assist algorithms (i.e. alignment or normalization process). Furthermore, the local feature based approaches trend to be extracted and computed directly on triangle mesh models rather than range images or point-clouds. Even though calculation complexity towards the latter two formats of data is relatively lower, the 3D meshed models are continuous and piece-wise smooth, which make it easier to compute the intrinsic geometric attributes and to design the associated feature descriptors.

Meanwhile, the preprocessing step also plays an important role in 3D face recognition, which needs to be explained and presented briefly here. Since 3D data conveys more useful information than 2D data, especially the shape-based information, the preprocessing using 3D data is more convenient and more accurate. Generally, in 3D face analysis, facial surface optimization and fiducial landmark detection are two main parts in preprocessing. The previous one contains the spike removing and hole filling. The fiducial landmarks are critical to estimate the initial positions of facial surfaces or locate more complex facial features (e.g. curve-based and region-based features). Curvature analysis is a dominant manner for this issue, and at present, considerable progress has been designed to ameliorate landmarking techniques. Many publications related to curvature analysis conclude that the detection accuracy of major landmarks, including nose tip, nose corners, inner eye corners, is almost 100% within an affordable error range in nearly frontal 3D face model [Colbry et al. 2005, Szeptycki et al. 2009, Zhao et al. 2009, Alyuz et al. 2010]. However, to precisely detect more landmarking points on faces under large pose variations still remains a problem. The ICP-based alignment algorithms, including the coarse to fine strategy, are effective for this problem, but they are also the most time-consuming process in 3D face recognition. The optimization of the preprocessing step in 3D face analysis is also needed to explore.
Chapter 2. Literature Review: 3D Face Recognition

2.5 Summary

In this chapter, we have extensively reviewed the approaches in three categories: holistic feature based approaches, local region-wise feature based approaches, and local point-wise feature based approaches. For each category, the most representative approaches and their corresponding achievement have been presented. As we discussed above, among these approaches, local feature based methods has the potential to boost recognition performance of holistic feature based methods. The local feature based approaches are the most promising category to achieve high discriminative power and to handle expression, pose and occlusion variations.

To achieve high recognition performance, local feature based 3D face recognition approaches usually describe facial surface with geometric attributes, i.e. the Gaussian curvature, the mean curvature, the principal curvatures, etc. The local features based on these geometric attributes are robust and effective in 3D face recognition. In Chapter 3, we will present a set of geometric attributes which are the curvatures generalized especially for 3D meshed surface, and introduce a local geometry-texture descriptor based method based on these generalized curvatures in Chapter 4.

Furthermore, as presented and discussed in this section, the normalization quality is a dominant factor in the local geometry-texture descriptor based method. Particularly, the head pose changes affect massively the comparison and the matching between two local point-wise feature. In Chapter 5, we introduced a registration-free local geometry-shape descriptor based on the same generalized curvatures. Due to the estimation of curvatures and the computation of feature are performed directly on triangle mesh, this method is a promising 3D face recognition algorithm to handle pose changes and external occlusions. Moreover, a further cross-resolution recognition is examined with this descriptor.

The general objective of all the presented approaches in this section is to improve the recognition performance and to handle the various environmental variations involved. However, the potential threat from spoofing attack also tests the robustness and the reliability of face recognition system. In chapter 6, we test the anti-spoofing performance of the local geometry-shape descriptor presented in chapter 5.
3.1 Introduction

With the development of 3D scanning and processing techniques, 3D faces are stored in point clouds, depth images, or triangle meshes. Point clouds and depth images are discontinuous, triangle meshes are only continuous and piece-wise smooth. Thus, in the context of face recognition, 3D face surfaces are generally non-differentiable discrete surfaces. Classical differential geometry is not adapted to study the shape of these surfaces, since the classical curvature formulas are adapted to (at least $C^2$) smooth surfaces. That is why to study the geometry of a 3D face and more generally the geometry of a triangle mesh, one needs to use other tools. Roughly speaking, there are two categories of curvature estimation methods on triangle mesh: discrete and continuous ones [Razdan & Bae 2005, Surazhsky et al. 2003]. All these methods estimate point-wise curvatures.
3. Reminders on Principal Curvature Measures

- The **discrete** methods estimate curvatures with operators acting directly on the triangle mesh. For example, Meyer *et al.* [Meyer *et al.* 2003] proposed to use the average Voronoi cells and mixed Finite-Element/Finite Volume method to define differential operators. Taubin [Taubin 1995] proposed to define the principal curvatures and principal directions from a $3 \times 3$ symmetric matrix using integral formulas. Berkmann [Berkmann & Caelli 1994] proposes to define the second fundamental form with covariance matrix and to further estimate the normal and the curvature measures of the points in point clouds.

- The **continuous** methods prefer to fit a smooth surface on or around the vertices of the triangle mesh, and estimate curvatures of the approximate smooth surface. Hamann [Hamann 1993] proposed to fit a quadric, Goldfeather *et al.* [Goldfeather & Interrante 2004] used a cubic. Theisel *et al.* [Theisel *et al.* 2004] proposed to estimate a curvature tensor through a vector field normal to each triangle facet. Rusinkiewicz [Rusinkiewicz 2004] used the robust statistics, i.e. M-estimation, to estimation both curvatures and their derivatives based on normal vectors.

In this thesis, we introduce a different approach to define curvatures on discrete surfaces. We introduce the concept of *curvature measure*, developed by [Cohen-Steiner & Morvan 2003] and based on geometric measure theory. Instead of computing the curvatures at a given point, one computes the curvatures of a Borel set (that is, we measure the curvature locally around the point). This allows to get interesting convergence and approximation theorems, when a sequence of meshes converges to a smooth surface. [Morvan 2008, Li *et al.* 2014b]. The following paragraphs summarize this construction, define the curvature measures in a general setting and in the special case of triangle meshes, that is, meshes adapted to our 3D face recognition problem.

### 3.2 Background on currents

We denote by $\mathbb{R}^n$ the $n$-dimensional $\mathbb{R}$-vector space, endowed with its standard scalar product $\langle \cdot, \cdot \rangle$. Details on basic definitions and properties on *differential forms* defined on $\mathbb{R}^n$ or a submanifold of $\mathbb{R}^n$ can be found in [Spivak 1981, Spivak 1965].

#### 3.2.1 General currents

We denote by $\mathcal{D}^m(\mathbb{R}^n)$ the $\mathbb{R}$-vector space of smooth *differential m-forms* with compact support on $\mathbb{R}^n$. It is endowed with the topology of uniform convergence on any compact subset of all partial derivatives of any order. We denote by $\mathcal{D}_m(\mathbb{R}^n)$
Chapter 3. Reminders on Principal Curvature Measures

the space of \emph{m-currents} of \( \mathbb{R}^n \), that is, the topological dual of \( \mathcal{D}^m(\mathbb{R}^n) \). The duality bracket will also be denoted by \( \langle \cdot, \cdot \rangle \). The \emph{support} \( \text{spt}(T) \) of a \( m \)-current \( T \) of \( \mathbb{R}^n \) is the smallest closed subset \( C \subset \mathbb{R}^n \) such that, if \( \omega \in \mathcal{D}^m(\mathbb{R}^n) \) satisfies \( \text{spt}(\omega) \cap C = \emptyset \), then \( \langle T, \omega \rangle = 0 \). We endow \( \mathcal{D}_m(TM^N) \) with the weak topology : if \( (T_k)_{k \in \mathbb{N}} \) is a sequence of \( m \)-currents of \( \mathbb{R}^n \) and if \( T \) is an \( m \)-current of \( \mathbb{R}^n \), then

\[
\lim_{k \to \infty} T_k = T \iff \forall \omega \in \mathcal{D}^m(\mathbb{R}^n), \lim_{k \to \infty} \langle T_k, \omega \rangle = \langle T, \omega \rangle.
\]  

(3.1)

The \emph{boundary} of a \( m \)-current \( T \) is the \( m-1 \) current \( \partial T \) defined for every \( \varphi \in \mathcal{D}^{m-1}(\mathbb{R}^n) \) by

\[
\langle \partial T, \varphi \rangle = \langle T, d\varphi \rangle.
\]

A current \( T \) is \emph{closed} if \( \partial T = 0 \).

3.2.2 Rectifiable currents

The link between \emph{currents} and \emph{measures} can be done as follows : We denote by \( \mathcal{H}^m \) the Hausdorff measure on \( \mathbb{R}^n \). A \emph{m-rectifiable} subset \( A \) of \( \mathbb{R}^n \) is the image of a bounded subset of \( \mathbb{R}^m \) by a Lipschitz map. It is well known that if \( A \) is \( m \)-rectifiable, then \( A \) admits a \( m \)-dimensional tangent vector space \( \mathcal{H} \)-almost everywhere, and one can assign to each of these spaces a unit \( m \)-vector \( v \) that orients \( A \). A \( m \)-current \( T \) with compact support in \( \mathbb{R}^n \) is \emph{rectifiable} if there exists a \( \mathcal{H}^m(\mathbb{R}^n) \)-measurable and \( m \)-rectifiable subset \( A \), an orientation defined at every point of \( A \) admitting a \( m \)-tangent vector space defined by a \( m \)-vector \( v \), an integrable function \( c \) with positive integer values defined at each point of \( A \) admitting a tangent vector space satisfying

\[
\int_A c \mathcal{H}^m < \infty,
\]

such that, for every \( m \)-differential form with compact support \( \omega \),

\[
\langle T, \omega \rangle = \int_A \langle v, \omega \rangle c \mathcal{H}^n.
\]

If \( B \) is any measurable set, the restriction of \( T \) to \( B \) is defined as follows :

\[
\langle T \upharpoonright B, \varphi \rangle = \int_A \langle v, \omega \rangle c \mathcal{H}^n.
\]

where \( \chi_B \) is the characteristic function of \( B \). A \( m \)-current is \emph{integral} if it is rectifiable and its boundary is rectifiable.

3.3 The normal cycle

Let \( T\mathbb{R}^3 \) denotes the tangent bundle of \( \mathbb{R}^3 \), identified to \( \mathbb{R}^3 \times \mathbb{S}^2 \). The unit tangent bundle \( \mathcal{S} \mathbb{R}^3 \subset T\mathbb{R}^3 \) of \( \mathbb{R}^3 \) can then be identified to \( \mathbb{R}^3 \times \mathbb{S}^2 \), where \( \mathbb{S}^2 \) denotes the unit sphere of \( \mathbb{R}^3 \). When it exists, the \emph{normal cycle} associated to a (compact
Chapter 3. Reminders on Principal Curvature Measures

singular) subset \( W \) of \( \mathbb{R}^3 \) is a closed integral 2-current \( N(W) \in D_2(T\mathbb{R}^3) = D_2(\mathbb{R}^6) \). It is the direct generalization of the unit normal bundle of a smooth surface of \( \mathbb{R}^3 \). A formal definition has been given in [Fu 1988]. This normal cycle can be defined for a large class of compact subsets, as convex subsets and polyhedra, in particular for triangle meshes. (But up to now, one does not know if this definition can be extended to associate a normal cycle to any compact subset of \( \mathbb{R}^6 \).) A compact subset \( G \) of \( \mathbb{R}^3 \) such that \( N(G) \) exists is said to be geometric, and \( N(G) \) is called its normal cycle. The main property of the normal cycle is its additivity [Fu 1988]:

**Proposition 1** If \( G_1 \) and \( G_2 \) are geometric, then \( G_1 \cup G_2 \) and \( G_1 \cap G_2 \) are geometric and

\[
N(G_1 \cup G_2) = N(G_1) + N(G_2) - N(G_1 \cap G_2).
\]

This inclusion-exclusion formula allows to compute explicitly the normal cycle of a geometric set \( G \) by decomposing the \( G \) in suitable simple subsets and computing the normal cycle of each subset. The basic examples of normal cycles are the following ones:

1. If \( D \) is a compact domain whose boundary is a smooth (oriented) surface \( S \) of \( \mathbb{R}^3 \), then its normal cycle is the closed current associated to its unit oriented normal bundle. (If there is no possible confusion, it is denoted by \( N(S) \) instead of \( N(D) \).)

2. If \( C \) is a convex body, then its normal cycle is the closed current associated to the oriented set

\[
\{(m, \xi) : m \in \partial C, \xi \in \mathbb{R}^3, ||\xi|| = 1, \forall z \in C, <\xi, \overline{mz}> \leq 0\}.
\]

3. The normal cycle of a (compact) domain \( W \) of \( \mathbb{R}^3 \) whose boundary is a polyhedron \( P \) can be computed by applying 3.2 to a decomposition of \( W \) into (convex) simplices, and using 2. (If there is no possible confusion, it is also denoted by \( N(P) \) instead of \( N(W) \).)

### 3.4 Curvature functions, curvature measures for smooth surfaces

In this section, we first review the classical pointwise second fundamental form, principal curvatures, principal vector fields, of smooth surfaces \( S \) embedded in \( \mathbb{R}^3 \). We introduce the concept of second fundamental measure and principal curvature measures for such smooth surfaces. For simplicity, we can assume that \( S \) bounds a compact domain \( D \) (and then has no boundary). This classical background can be found in [Chen 1973, Morvan 2008, do Carmo Valero 1992] for instance.
3.4.1 Classical principal curvatures of smooth surfaces in $\mathbb{E}^3$

Let us consider a smooth oriented surface $S$ embedded in the Euclidean space $\mathbb{E}^3$. Let $\xi$ be a normal vector field, $h$ the second fundamental form of $S$ in the normal direction $\xi$. At each point $p$ of $S$, the eigenvalues of $h_p$ are called the principal curvatures $\lambda_1 p$, $\lambda_2 p$ of $S$. The eigenvectors of $h_p$ (tangent to $S$) are called the principal curvature vectors $e_1 p$, $e_2 p$ at $p$. In an orthonormal frame of eigenvectors $(e_1 p, e_2 p)$ at $p$, the matrix of $h_p$ is

$$
\begin{pmatrix}
\lambda_1 p & 0 \\
0 & \lambda_2 p
\end{pmatrix}.
$$

(3.3)

The local bending informations (around any point $p$) of $S$ can be described by its second fundamental form, in particular by its principal curvatures and principal directions.

One remarks that the previous constructions and definitions make sense because of the smoothness of $S$: Indeed, to define $h$, one needs to differentiate a parametrisation of $S$. To describe the local bending information of triangle meshes, one must generalize these constructions to non smooth objects. A possible solution is to replace functions by measures. It is the goal of section 3.4.2 in the smooth surface case: We will associate a measure to $h$, and in section 3.6, we will define the corresponding measure for any triangle mesh.

3.4.2 Second fundamental measures on smooth surfaces of smooth surfaces in $\mathbb{E}^3$

Let $S$ be a smooth surface (to simplify, we can assume that it bounds a compact domain, and we denote by $\xi$ the outward normal vector field). To any Borel subset $B$ of $\mathbb{R}^3$ and any vector field $X$ and $Y$ of $\mathbb{E}^3$, we define

$$
\overline{h}_B(X,Y) = \int_{S \cap B} h_p(pr_{T_pS}X_p, pr_{T_pS}Y_p) dp,
$$

(3.4)

where $pr_{T_pS}$ denotes the orthogonal projection over the tangent plane $T_pS$ of $S$ at $p$. For fixed $X$ and $Y$, the map

$$
B \to \overline{h}_B(X,Y)
$$

(3.5)

is a signed measure on $\mathbb{R}^3$. Moreover, if we restrict the set of vector fields $X$ and $Y$ in $\mathbb{E}^3$ to constant vector fields in $\mathbb{R}^3$, then for any fixed Borel subset $B$ of $\mathbb{E}^3$, the map :

$$
\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R},
$$

(3.6)

\[(X,Y) \to \overline{h}_B(X,Y)\]
Chapter 3. Reminders on Principal Curvature Measures

is a symmetric bilinear form on $\mathbb{R}^3$. Let $(\lambda_{1B}, \lambda_{2B}, \lambda_{3B})$ denotes its three eigenvalues and $(e_{1B}, e_{2B}, e_{3B})$ the corresponding eigenvectors. For every $i \in \{1, 2, 3\}$, the map

$$\lambda_i : B \rightarrow \lambda_{iB}$$

(3.7)

is a measure on $\mathbb{R}^3$ called the $i^{th}$ principal curvature measure. In the same way, the eigenvectors of $\overline{h}_B$ are called the principal curvature vectors of $S$ over $B$. Remark that we get now three principal curvatures instead of two in the pointwise approach. In the orthonormal frame of principal vectors, the matrix of $\overline{h}_B$ is now

$$
\begin{pmatrix}
\lambda_{1B} & 0 & 0 \\
0 & \lambda_{2B} & 0 \\
0 & 0 & \lambda_{3B}
\end{pmatrix}.
$$

(3.8)

Moreover, let us remark that if we suppose that the Borel subset $B$ is reduced to a point $p \in S$, and $y_p \in T_pS$, $\overline{h}(p)$ can be considered as a Dirac measure, with $\overline{h}_S^x\{p\} = h_p(y_p, y_p)$, where $T_pS$ denotes the tangent plane of $S$ at $p$. If $z_p = \xi_p$, where $\xi_p$ denotes the normal vector of $p$, then $\overline{h}_S^x\{p\} = 0$. Then, in the frame of $(e_{1p}, e_{2p}, \xi_p)$ the matrix of $\overline{h}$ becomes to

$$
\begin{pmatrix}
\lambda_{1p} & 0 & 0 \\
0 & \lambda_{2p} & 0 \\
0 & 0 & 0
\end{pmatrix}.
$$

As mentioned in [Cohen-Steiner et al. 2006], [Sun & Morvan 2015, Morvan 2008], because our purpose in 3D face recognition will use the triangle mesh, it appears useful to introduce a slightly different measure: We define

$$\overline{h}_B^S(X, Y) = \int_{S \cap B} h_p(j \circ pr_{T_pS}X_p, j \circ pr_{T_pS}Y_p)dp,$$

where, at each point $p$, $j$ is the direct rotation of angle $\frac{\pi}{2}$ in the tangent plane $T_pS$. Consequently, at each point $p$, one must swap the eigenvalues in the tangent plane, that is, replace the matrix in (3.8) by the matrix $\begin{pmatrix}
\lambda_{1B}^S & 0 & 0 \\
0 & \lambda_{2B}^S & 0 \\
0 & 0 & \lambda_{3B}^S
\end{pmatrix}$. The theoretical approach is exactly the same and the computations are simpler.
Chapter 3. Reminders on Principal Curvature Measures

3.5 Curvature measures for singular spaces

The goal of this section is to define a framework that gives the possibility to define \textit{curvature measures} for a large class of singular objects, namely (in our context), geometric subsets of \( \mathbb{R}^3 \), and that generalize the curvature measures defined in section 3.4.2 for smooth objects. The idea is to define such measures by means of a differential forms on the unit tangent bundle of \( \mathbb{R}^3 \).

We denote by \( ST \mathbb{R}^3 \subset T \mathbb{R}^3 \) the bundle of unit tangent vectors of \( \mathbb{R}^3 \) (identified to \( \mathbb{R}^3 \times S^2 \)), where \( S^2 \) denotes the unit sphere of \( \mathbb{R}^3 \). For every vector fields \( X \) and \( Y \) on \( \mathbb{R}^3 \), the \textit{second fundamental curvature form} is the 2-differential form \( \omega^{X,Y} \) on \( ST \mathbb{R}^3 \) defined as follows: at each point \( (m, \xi) \in ST \mathbb{R}^3 \),

\[
\omega^{X,Y}_{m,\xi} = X \times (\xi \times Y),
\]

where \( \times \) denotes the vector product in \( \mathbb{R}^3 \). We can now define the \textit{second fundamental curvature measure} of any geometric set:

\textbf{Definition 1} Let \( G \) be a geometric subset of \( \mathbb{R}^3 \). The \textit{second fundamental vector valued measure} \( \Phi_G \) of \( G \) is defined as follows: For any vector fields \( X \) and \( Y \) on \( \mathbb{R}^3 \),

\[
\forall B \in B(\mathbb{R}^3), \Phi_G^{X,Y}(B) = N(G)_{\nu}(B \times \mathbb{R}^3)\omega^{X,Y}.
\]

(3.9)

The crucial point is the following result [Cohen-Steiner et al. 2006, Cohen-Steiner & Morvan 2003]:

\textbf{Theorem 1} If \( G \) is a smooth surface, then \( \Phi_G = \overline{\mathbf{h}} \).

This theorem claim that if the geometric set \( G \) is a smooth surface, then \( \Phi_G \) is nothing but "the integral of \( \mathbf{h} \) over \( G \)." Thus, \( \Phi_G \) is a "good" generalisation of \( \mathbf{h} \). Mimicking Section 3.4.2, we now restrict the set of vector fields \( X, Y \) in \( \mathbb{R}^3 \) to \textit{constant} vector fields. Then, for any fixed Borel subset \( B \) of \( \mathbb{R}^3 \), the map:

\[
\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R},
\]

\[
(X, Y) \to \Phi_G^{X,Y}(B)
\]

(3.10)

is a bilinear symmetric form on \( \mathbb{R}^3 \). Let us denote by \( (\lambda_1^G, \lambda_2^G, \lambda_3^G) \) its three eigenvalues. For every \( i \in \{1, 2, 3\} \), the map

\[
\lambda_i^G : B \to \lambda_i^G
\]

(3.11)

is again a \textit{measure} on \( \mathbb{R}^3 \) called the \textit{i}th \textit{principal curvature measure} of \( G \) over \( B \). In the same way, the corresponding eigenvectors are called the \textit{principal curvature vectors} of \( G \) over \( B \). In the orthonormal frame of principal vectors, the corresponding matrix is
Chapter 3. Reminders on Principal Curvature Measures

\[
\begin{pmatrix}
\lambda_1^G & 0 & 0 \\
0 & \lambda_2^G & 0 \\
0 & 0 & \lambda_3^G
\end{pmatrix}.
\]  

(3.12)

3.6 Principal curvature measures of triangle meshes in \( \mathbb{E}^3 \)

We now deal with triangular meshes \( T \) in \( \mathbb{E}^3 \). For simplicity, as for smooth surfaces, we assume that \( T \) bounds a compact domain. Since a triangular mesh is a singular surface, one cannot describe its shape by using a point-wise approach. That is why we will use the measure theoretic approach described in Section 3.5. We denote by \( E \) the set of edges \( e \) of \( T \), \( l(e \cap B) \) the length of \( e \cap B \), and \( \angle(e) \) the signed angle between the unit oriented normal \( n_1 \) of face \( f_1 \) (resp. \( n_2 \) of face \( f_2 \)) incident to \( e \) (its sign is positive if \( e \) is convex and negative otherwise). Fig.3.1 shows an example of \( \angle(e) \). For any Borel subset \( B \) of \( \mathbb{R}^3 \), and any constant vector fields \( X \) and \( Y \) in \( \mathbb{R}^3 \), we deduce from 3.9 by a direct computation the following explicit formula :

\[
\Phi_T^{XY}(B) = \sum_{e \in E} l(e \cap B) \angle(e) \langle X, e \rangle < Y, e \rangle.
\]  

(3.13)

In other words, in any fixed frame, the matrix \( M_T^B \) associated to \( h_B \) satisfies :

\[
M_T^B = \sum_{e \in E} l(e \cap B) \angle(e) e \cdot e'.
\]  

(3.14)

A classical explicit diagonalization of \( M_T^B \) gives the three principal curvature measures of \( B \) and the corresponding diagonalised matrix

\[
\begin{pmatrix}
\lambda_1^T & 0 & 0 \\
0 & \lambda_2^T & 0 \\
0 & 0 & \lambda_3^T
\end{pmatrix}.
\]  

(3.15)

Similarly to the smooth case, the set of eigenvalues \( \{ \lambda_1^B, \lambda_2^B, \lambda_3^B \} \) of \( h_B \) are called the principal curvature measures of \( T \) over \( B \) and the eigenvectors \( \{ e_1^B, e_2^B, e_3^B \} \) of \( h_B \) are called the principal curvature vectors of \( T \) over \( B \).

3.7 A convergence theorem

To justify one more time the framework described in section 3.3, we now state the following convergence result [Sun & Morvan 2015, Morvan 2008]. Roughly speaking, it can be stated as follows : If a sequence \( T_k \) of triangular meshes of \( \mathbb{R}^3 \) tends
Chapter 3. Reminders on Principal Curvature Measures

Figure 3.1: The angle between two oriented triangles incident to $e$

(for a suitable topology) to a smooth surface $S$, then the corresponding second fundamental measures $\Phi_T$, (resp. eigenvalues $\{\lambda^T_i, (1 \leq i \leq 3)\}$), tend to $\Phi_S$, (resp. $\{\lambda^S_i, (1 \leq i \leq 3)\}$). To be accurate, we will use the following terminology [Morvan 2008, Morvan & Thibert 2004]:

- The fatness $\Theta(T)$ of a triangular mesh $T$ is defined as follows: If $t$ is a triangle, we begin to define the size $\varepsilon(t)$ of $t$: it is the maximum of the length of its edges $e$. Moreover, the fatness of $t$ is the real number

  $$\Theta(t) = \frac{\text{area}(t)}{\varepsilon(t)^2}.$$  

  Finally, the fatness of $T$ is the minimum of the fatness of its triangles. We denote by $F_\theta$ the class of triangular meshes in $\mathbb{E}^3$ with fatness greater or equal to $\theta$.

- A triangular mesh $T$ in $\mathbb{R}^3$ is closely inscribed in a smooth surface $S$ if its vertices belong to $S$ and if the orthogonal projection of $T$ onto $S$ is a bijection.

It is well known that in general, the Haussdorf convergence of a sequence of triangular mesh to a smooth surface does not imply the convergence of their geometric invariant (as shown by the classical lantern of Schwarz example [Morvan 2008, Morvan & Thibert 2004]). However, adding assumptions on the fatness of the triangular meshes, one can prove the following result in [Morvan 2008, Cohen-Steiner & Morvan 2003, Cohen-Steiner et al. 2006]:

**Theorem 2** Let $S$ be an (oriented closed) smooth surface of $\mathbb{R}^3$. Let $T_k$ be a sequence of (oriented closed) triangular meshes $T_k$ closely inscribed in $S$ such that:

1. The limit of $T_k$ is $S$ for the Hausdorff distance,
Chapter 3. Reminders on Principal Curvature Measures

2. the fatness of $T_k$ is uniformly bounded by below by a positive constant: there exists $\theta > 0$ such that for all $k \in \mathbb{N}$, $T_k \in F_\theta$.

Then,

1. the sequence of second fundamental measures $\Phi_{T_k}$ weakly converges to $\Phi_S = \overline{h}^S$.

2. The principal curvature measures $\lambda_i^{T_k}$, weakly converge to the curvature measures $\lambda_i^S$, $(1 \leq i \leq 3)$.

In particular, for any Borel subset $B$ whose boundary is empty, and fixed $X$ and $Y$, the sequence $\Phi^{X,Y}_{T_k}(B)$ converges to $\Phi^{X,Y}_S(B)$, and the principal curvature measures $\lambda_i^{T_k}(B)$, converge to the curvature measures $\lambda_i^S(B)$.

After all, if the triangle mesh converges to a smooth surface, two principal curvature measures converge to two principal curvatures and the third one converges to 0. Meanwhile, in the same case, two principal curvature vectors converge to the principal directions and the third one converges to the normal direction.

3.8 Conclusion

In this chapter, we introduce a group of geometric attributes, the principal curvature measures, associated to a triangle mesh. They generalize the integral of the principal curvatures of smooth surfaces. In the following chapters, we will adopt these principal curvature measures in the design of 3D facial feature descriptors. The generalized principal curvature measures have the following threefold advantages in shape-analysis based 3D face recognition: (1) They can be defined and computed directly on triangle meshes. This makes the curvature and direction information of a triangle mesh more reasonable and reliable. Meanwhile, these geometric attribute represent the shape-based information of original 3D meshed facial surface, rather than the shape of the manufactured surface fitting to the original 3D meshed face. (2) In contrast to other meshed surface based curvature estimation, three principal curvature measures are achieved reasonably instead of two. Particularly, the third principle curvature measure $(\lambda_3)$ enhances the facial shape based description ability. We will show that the three principal curvature measures are useful to describe the geometric structure of a 3D face. Moreover, they offer crossing complementary information for 3D face recognition. That is, different principal curvature measures highlight different and complementary meaningful areas on 3D face. (3) Since the generalized second fundamental measure depends on the scale of $B$, the principal curvature measures are also affected by the scale. This means that the various scales of $B$ can fit different 3d meshed face models with different scales. In consideration of these advantages, we will propose two 3D facial feature descriptor based
Chapter 3. Reminders on Principal Curvature Measures

on principal curvature measures, and evaluate their discriminative power on the comprehensive experiments.

However, we also need to notice that the calculation of these principal curvature measures highly depend on the topology structure of triangle mesh. It indicates that the change of the length of edge and the change of the angle between edges have impact on the calculation results. Moreover, the calculation method is based on the Borel subset around each point. Therefore, the calculation results also depend on the choice of Borel subset. According to these two conclusions, in the 3D face recognition, it is better to extract the feature associated to a local region around a point to enhance the robustness of the feature. Meanwhile, it is worthy to evaluate the impact of Borel subset on the performance of the proposed methods.
Chapter 4

3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Contents

4.1 Introduction ................................................. 67
4.2 Local Principal Curvature Measures Pattern (LPCMP) feature descriptor ............................................. 68
  4.2.1 Generating Curvature Faces ................................. 69
  4.2.2 LPCMP feature descriptor extracted in curvature faces .................................................. 71
  4.2.3 Sparse Representation-based Classifier and Fusion .................................................. 72
4.3 Experiments .......................................................... 73
  4.3.1 Database ...................................................... 73
  4.3.2 Experiment Settings ............................................ 74
  4.3.3 Experiment Results ............................................ 75
4.4 Conclusion ............................................................ 82

4.1 Introduction

With the development of 3D scanning and processing techniques, 3D face recognition have been widely studied. Especially, since 3D face sensors can accurately and sensitively capture the geometrical shape of the underlying 3D facial surfaces, designing a discriminating facial geometric surface feature is a critical issue in 3D face recognition. In general, the normal and the curvatures are the most commonly used geometric features to describe the facial local surface. Maes et al. applied the mean curvature in DoG (Difference of Gaussian) based scale space to detect salient vertices and then adopted histogram of shape index (calculated with the maximal curvature and the minimal curvature) in local regions to build the descriptor [Maes et al. 2010]. Kakadiaris et al. analyzed the normal map and geometry image by using a wavelet transform [Kakadiaris et al. 2007]. Szeptycki et al. adopted the
mean curvature and the Gaussian curvature to locate the most salient facial feature points (e.g., nose tip and two eye inner corners) [Szeptycki et al. 2009]. Li et al. proposed Multiple salient vertices based Histograms of Multiple order surface differential Quantities computed in the neighborhood of salient vertices automatically using the histograms of shape index and gradients of shape index based on maximal and minimal curvatures through a 3D Gaussian scale space [Li et al. 2015]. Tonchev et al. processed the curvature analysis and range image representation on the input point cloud [Tonchev et al. 2013]. Hwang et al. extend Gabor wavelet kernels by adding a spatial curvature term and adjust the width of the Gaussian at the kernel for a low-resolution image [Hwang et al. 2011]. These works proposed efficient facial surface features and achieved outstanding experiment results.

In this chapter, we propose a geometry-texture based 3D facial feature descriptor involving the principal curvature measures presented in Chapter 3. Principal curvature measures are calculated on each vertex in 3D meshed face scan, and three curvature faces are generated by mapping the vertex-related principal curvature measures in 2D images. The feature is extracted by encoding the curvature faces and compute the histograms of the mini-patches in the encoded curvature faces. We name this 3D face recognition framework as Local Principal Curvature Measure Pattern (LPCMP) based 3D FR method. The identification and verification performance of this method are evaluated on the FRGC v2.0 database.

4.2 Local Principal Curvature Measures Pattern (LPCMP) feature descriptor

As the presentation of principal curvature measures in Chapter 3, these geometric attributes can handle with the description of the shape information of meshed surface, e.g. 3D meshed face surface. In order to characterize the facial surface by principal curvature measures, we introduce the Local Principal Curvature Measure Pattern (LPCMP) feature descriptor. It computes these geometric attributes directly on 3D meshed faces and generates the LBP-like facial feature based on them. Here, we choose LBP because of its efficient encoding pattern but effective performance in related face recognition researches [Li et al. 2014a, Huang et al. 2010]. The feature matching step is performed with Sparse Representation Classifier (SRC) based matching method, which has been proven discriminating in face processing [Wright et al. 2009, Li et al. 2014a]. The framework of this method is shown in Fig.4.1, which consists of the following components: face preprocessing, face alignment, principal curvature measures estimation, curvature faces generation, feature descriptor extraction, and SRC-based keypoint matching and fusion. The details of these processing steps are introduced in the following sections.
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Figure 4.1: The framework of 3D face recognition based on LPCMP-based feature descriptor.

4.2.1 Generating Curvature Faces

Due to the stability and discriminative power of Local Binary Pattern (LBP) algorithm, we combine the principal curvature measures and LBP encoding method to generate the LPCMP feature descriptor. The basic idea of this feature is to generate a principal curvature measures related 2D texture image of face, and to compute the LBP value and the feature on these texture images. Here, we name the textures image with principal curvature measures as curvature faces. Obviously, three curvature faces are produced corresponding to three principal curvature measures.

For guaranteeing the quality of LPCMP feature descriptor, the face is firstly normalized and aligned to frontal view. We choose three landmarks (e.g. two eye inner corners and nose tip) for normalization and alignment. The location of landmarks is confirmed following Szeptycki et al. work in [Szeptycki et al. 2009]. The scale of curvature faces is normalized according to the distance between two eye corners and the distance between the nose tip and the midpoint of two eye corners. The alignment of face is performed based on the symmetry plane [Wang et al. 2010b] of face, and the angle between the $xy$ plane and the line connecting the nose tip and the midpoint of two eye corners. After these two steps, the face model with principal curvature measures is projected to $xy$-plane to generate a 2D image in resolution of $240 \times 200$. Specifically, we find the mapping between the points in 3D meshed face model and the pixel in 2D face image. The value of the pixel is the principal curvature measure of the corresponding points. If the projected vertex locates on the pixel in this 2D image, the value of this corresponding pixel is the principal curvature measure. Otherwise, we apply the area based interpolation to estimate the value of the pixel as follows. This 2D image associated to one principal curvature measure is named as curvature face.

The area based interpolation work like this: Here we mean to determine the
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Figure 4.2: Illustration of the principal curvature measures computed on a 3D meshed face scan (A), and the corresponding curvature faces, the projected principal curvature measures in 2D image (B).

value of the pixel \( p \) regarded as the target pixel, but its corresponding point \( P \) in the triangle mesh does not belong to the set of vertices. The point \( P \) locates in a triangle facet \( \triangle ABC \) (\( A, B \) and \( C \) are the vertices of mesh model). Thus, the principal curvature measures of \( P \) is estimated as the area interpolation method defined as:

\[
C_P = \frac{S_{\triangle PBC}}{S_{\triangle ABC}} C_A + \frac{S_{\triangle PCA}}{S_{\triangle ABC}} C_B + \frac{S_{\triangle PAB}}{S_{\triangle ABC}} C_C \tag{4.1}
\]

where \( C_P, C_A, C_B \) and \( C_C \) are the principal curvature curvatures of points \( P, A, B \) and \( C \), and \( S \) is the triangle area. Then the value of the pixel \( p \) is the principal curvature measure of \( P \), that is, \( C_P \).

The samples of three curvature faces are shown in Fig.4.2(B). This figure shows clearly that each curvature faces generated by corresponding principal curvature contain more informative geometric information than their corresponding range image which looks quite smooth. Especially, the geometric shape details around the eye corners, mouth corners and cheek region are quite well highlighted.

70
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

4.2.2 LPCMP feature descriptor extracted in curvature faces

Once three curvature faces in mesh model are estimated, we encode the curvature faces following LBP-based method. The code of each pixel in the curvature component face is generated based on its local neighbourhood predefined. The neighbourhood is defined as a set of sampling points forming a circle centered at the encoded pixel. If the sampling points do not locate at the pixel in the curvature component face, we estimate its value with bilinear interpolation. This encoding method is usually classified with $Q_{r;p}$, which denotes the parameters of this coding that $r$ is the radius of the circle, and $p$ is the amount of sampling points. Fig. 4.3 shows two examples of neighbourhood of the central point for LBP.

Then we compare the value of target point $p_c$ and each sampling point $p_s$ followed in the rule that $\lambda^k_{p_c} \leq \lambda^k_{p_s}$ encodes it with 0 and $\lambda^k_{p_c} > \lambda^k_{p_s}$ encodes it with 1. Here, $\lambda^k$ is the $k$-th curvature face corresponding to $k$-th principal curvature measure. Connecting these binary codes starting from the top-left in the clockwise direction, we obtain the binary number, and its corresponding decimal number is set as the updated value of target point in LPCMP-related image. The encoded curvature faces are shown in Fig. 4.4. Formally, the decimal number of the target central point $p_c$ is:

$$LPCMP(Q_{r;p}(p_c)) = \sum_{s=1}^{p} t(\lambda^k_{p_c} - \lambda^k_{p_s}) \times 2^s, k \in 1, 2, 3$$

where $t(x) = 1$, if $x \leq 0$ and $t(x) = 0$, if $x > 0$.

In order to characterize the shape of a local region rather than a point, histogram-based statistics are computed and used as a facial feature vector in a series of separated patches. In our approach, the encoded curvature face is sepa-
Rated in 12 × 10 patches. In each patch, the histogram of encoded curvature face is computed for each principal curvature measure, and histograms are concatenated as the global facial feature for utilizing the spatial information of 3D facial surface. Three global facial features are generated respectively corresponding to three principal curvature measures.

4.2.3 Sparse Representation-based Classifier and Fusion

Sparse representation-based classifier applied in this paper is following the presentation of Li et al.’s work in [Li et al. 2014a] which demonstrated in particular the effectiveness of Sparse Representation-based Classifier (SRC) in 3D face recognition and it can be solved by the OMP algorithm [Pati et al. 1993]. The minimal reconstruction residual $r_i$, as the matching score calculated by SRC, can be used to determine the similarity between two face.

Specifically, given a query 3D face model, three global histogram-based LPCMP features are computed. Then, based on the gallery dictionary, three sets of reconstruction residual are achieved. For each set of reconstruction residual, the identity of the query face is label as the one of the gallery sample if the corresponding reconstruction residual error is the smallest one. Furthermore, we take three score-level fusion rules to fusing the reconstruction residual of each principal curvature mea-
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

...sure, including minimum fusion, mean fusion and dynamic fusion [Mian et al. 2008]. The similarity score vector of one probe sample to all gallery samples is normalized to the interval of [0,1] using the min-max rule before the score-level fusion. The minimum fusion is to select the minimum value in the candidate reconstruction residual as the final similarity measurement. The mean fusion is to compute the mean of all candidate reconstruction residual as the fused reconstruction residual. The dynamic fusion is defined as:

$$ R = \sum_{i=1}^{3} w_{\lambda_i} \times r_{\lambda_i} $$

(4.3)

where $r_{\lambda_i}$ denotes the reconstruction residual error vector of $i$-th principal curvature measure based descriptor, and $w_{\lambda_i}$ denotes the corresponding weight defined as:

$$ w_{\lambda_i} = \frac{\text{mean}(r_{\lambda_i}) - \text{min}_1(r_{\lambda_i})}{\text{mean}(r_{\lambda_i}) - \text{min}_2(r_{\lambda_i})} $$

(4.4)

where the operators $\text{min}_1$ and $\text{min}_2$ denote the first and the second minimum value of reconstruction residual error vector. Similarly, the identity of the query face is labelled as the class of the gallery possessing the $\text{min}(R)$. In the following experiment, we will test the discriminative power of LPCMP feature generated by individual principal curvature measure and their fusion. The purpose of this setting is to demonstrate that three principal curvature measures can describe the shape of facial surface, and the complementary information carried by them is able to improve the performance of feature descriptor.

4.3 Experiments

In this section, we introduce the database and describe its particularities. Then we present our experiment settings and show the experiment results for examine the performance of LPCMP-based 3D face recognition system with three principal curvature measures. Specifically, we will test the impact of the region of Borel subset, the selection of parameters in curvature faces generation, the recognition performance of three individual principal curvature measure, and the three fusion schemes of three principal curvature measures.

4.3.1 Database

In the experiment part, we adopt the FRGCv2 database for testing the discriminative power of our proposed method. The FRGCv2 database [Phillips et al. 2005] is made up of 4007 textured 3D face scans of 466 subjects with different facial expressions. The face scans are captured in controlled lighting and pose by the Minolta...
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Figure 4.5: Examples of pre-processed 3D face scans in the FRGCv2 database.

Vivid 900 scanner. FRGCv2 database is the largest public 3D face database in the literature and has become the benchmark for evaluating 3D FR approaches. In the experiment, we take the earliest image of each person to group the gallery set (466 images) and the rest to form the probe gallery (3541 images). All scans in FRGCv2 database are preprocessed by using 3D Face Models Preprocessing Tool\(^1\). It is developed by Szeptycki et al. [Szeptycki et al. 2009]. The preprocessing work contains: spike and noise removing, holes filling, nose tip and eye corners localization and face cropping. As introduced in [Szeptycki et al. 2009], a decision-based median filtering technique is used to remove spikes. The holes are detected by searching vertices who have less than 8 adjacent vertices, and filled by fitting square surfaces. The nose tip and the eye corners are located through a curvature analysis-based coarse grained search and generic face model-based fine grained search. Some examples of the pre-processed samples are shown in Fig. 4.5.

4.3.2 Experiment Settings

To comprehensively evaluate the proposed method, four experiments are designed for four purposes.

1. To examine the impact of the size of Borel subset for principal curvature measures estimation, we test the rank-1 identification accuracy with three Borel subset in 3, 5 and 7-ring. The parameters utilized in curvature faces encoding is set as $Q_{2,16}$.

2. To evaluate the effectiveness of different encoding pattern used in LPCMP encoding step, we choose $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$ to generate the LPCMP value of each pixel. Here, the size of Borel subset is 5-ring.

3. To determine if the three principal curvature measures carry the complementary shape information, we evaluate the fusion method with the rank-one recognition rate for different combinations of the features: $\lambda_1 + \lambda_2$, $\lambda_2 + \lambda_3$, $\lambda_3 + \lambda_1$, and $\lambda_1 + \lambda_2 + \lambda_3$. Moreover, we test the performance of three score-level fusion

\(^1\)http://pszeptycki.com/tool.html
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Table 4.1: Rank-one recognition rate achieved on FRGCv2 database for evaluating the impact of region size of Borel subset.

<table>
<thead>
<tr>
<th>Borel subset size</th>
<th>3-ring</th>
<th>5-ring</th>
<th>7-ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>85.82%</td>
<td>85.87%</td>
<td>85.70%</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>85.87%</td>
<td>85.48%</td>
<td>85.33%</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>83.06%</td>
<td>85.39%</td>
<td>84.46%</td>
</tr>
</tbody>
</table>

method in this protocol: the minimum fusion rule, the mean fusion rule and the dynamic fusion rule.

4. To test the verification performance of the method, a related experimental scenario is performed on FRGCv2 database.

4.3.3 Experiment Results

4.3.3.1 Impact of region size of Borel subset

From the presentation of principal curvature measures in Chapter 3, if we mean to describe the shape information of a vertex on the meshed facial surface, we actually describe the shape information of the Borel subset associated to this vertex. The estimation of principal curvature measures is associated closely to the region size of Borel subset. Meanwhile, the region size of Borel subset also has impact on the facial feature and the recognition result. Motivated by this reason, in this part, we test the impact of region size of Borel subset to the rank-one recognition rate with three individual features.

In the experiments, the Borel subset is regarded as a sphere-like region associated to a vertex. Specifically, given a vertex $v$, we define the set of the $k$-ring neighbouring vertices around $v$ is the intersection region between the surface and the Borel subset. If the least number of edges needed to connect a vertex to $v$ is equal or smaller than $k$, this vertex is in this $k$-ring neighbourhood. Through adjusting the value of $k$, we control the estimation region of principal curvature measures of Borel subset associated to $v$. This settings will be adopted in all following experiments related to principal curvature measures.

In this part, the region sizes of Borel subset defined as 3, 5, and 7-ring are the testing candidates. We evaluate the rank-one recognition rate for the LPCMP-based feature descriptor based on three individual principal curvature measure. The neighbourhood pattern for encoding curvature faces is $Q_{2,16}$. There are 120 patches (i.e. $12 \times 10$ patch grid) for generating the local region-based histograms of curvature face.

Table 4.1 shows the experimental results evaluated on FRGCv2 database. Each
column contains the results corresponding to one size of Borel subset, and each row shows the recognition rates corresponding to one feature associated to one principal curvature measure. From the viewpoint of the performance of each principal curvature measure, the recognition rates based on each principal curvature measure are about 85%. The first conclusion achieved is that each of them is able to describe the facial surface, and their corresponding feature descriptor can quantize the facial surface and obtain satisfactory recognition performance.

However, if we compare the results between different sizes of Borel subset, the recognition rate seems not to be sensitive to the change of the size of Borel subset. For different cases, the recognition rates almost remain similar. There is only a minor change when using the feature generated by third principal curvature measure. It is hard to speak which size of Borel subset is most suitable for this approach. But from the presentation in Chapter 3, the region size of Borel subset affects intensively the estimated value of principal curvature measures. When we analyse our algorithm carefully, we find out the reason causing this phenomenon. The first reason is that we cannot define the estimation correctness of principal curvature measures in 3D face recognition case, because the ground-truth of principal curvatures of the facial surface is completely unknown. Therefore, from the achieved experimental results, it is not clear to decide which size of Borel subset is the best to estimate principal curvature measures. However, the distribution of principal curvature measures is similar when the size of Borel subset changes. Meanwhile, because the encoding method can be regarded to calculate the gradient of the pixel values (i.e. the individual principal curvature measure projected to 2D image), the similar distributions of principal curvature measures generate the similar curvature faces. The histogram-based feature extraction method further reduces the impact of the size of Borel subset in 3D face recognition.

4.3.3.2 The impact of encoding pattern

As described in feature extraction section, the encoding pattern \( Q_{r,p} \) consists of two parameters: \( r \) refers to the radius of encoding circle, and \( p \) refers to the number of neighbour points used for encoding. In this part, we learn the impact of different encoding patterns with various \( r \) and \( p \). Specifically, given a curvature face with one principal curvature measure estimated in a Borel subset of 5-ring size, we encoded the curvature face under three patterns: \( Q_{1,8} \), \( Q_{2,16} \) and \( Q_{3,24} \). Similar to the part above, we perform the identification experiments on three features related to three principal curvature measures estimated in Borel subset with geodesic radius 5 respectively. The experimental results are shown in Table 4.2.

In Table 4.2, each column shows the rank-one recognition rate of feature based on one encoding pattern, and each row displays the identification accuracy associ-
Table 4.2: Rank-one recognition rate achieved on FRGCv2 database for evaluating the impact of encoding pattern.

<table>
<thead>
<tr>
<th>Encoding Pattern</th>
<th>$Q_{1,8}$</th>
<th>$Q_{2,16}$</th>
<th>$Q_{3,24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{1B}$</td>
<td>83.61</td>
<td>85.87%</td>
<td>81.27%</td>
</tr>
<tr>
<td>$\lambda_{2B}$</td>
<td>84.86</td>
<td>85.48%</td>
<td>83.74%</td>
</tr>
<tr>
<td>$\lambda_{3B}$</td>
<td>83.75</td>
<td>85.39%</td>
<td>80.60%</td>
</tr>
</tbody>
</table>

ated to each principal curvature measure. Even though the recognition accuracy towards each principal curvature measure is over 80%, but the change of encoding pattern affects obviously the recognition performance of extracted feature. For each principal curvature based feature descriptor, the encoding pattern $Q_{2,16}$ performs better than the other two encoding patterns. $Q_{2,16}$ raises 1.2% of the recognition rate with $Q_{1,8}$ and 3.1% with $Q_{3,24}$. This encoding pattern makes it possible to capture enough details of a facial shape to efficiently discriminate the identity. Based on this experiment, we decide to choose $Q_{2,16}$ as the encoding pattern to encode the curvature face and generate the LPCMP feature descriptor.

4.3.3.3 Identification test with score-level fusion rules

From the two previous experiments, we have achieved three important conclusions: (1) the variation of region size of Borel subset has limited impact on the proposed 3D face recognition method; (2) the selection of encoding pattern has significant impact on the recognition performance of our method, and $Q_{2,16}$ is the best one of three examined encoding pattern; (3) three principal curvature measures can describe the shape of facial surface individually, and their related LPCMP feature can characterize the facial surface respectively. Based on these conclusions, we plan to evaluate the identification performance by fusing the principal curvature measures for testing their complementarity on shape-based information. In this part, three score-level fusion rules are performed to fusing three principal curvature measures related facial feature, including the minimum fusion, the mean fusion, and the dynamic fusion. Furthermore, even though the experimental results have demonstrated that the impact of region size of Borel subset is limited, it is also meaningful to test the performance by fusing different regions sizes of Borel subset to form a multi-scale facial feature descriptor. According to the experimental results achieved above, the encoding pattern is chosen as $Q_{2,16}$.

Table 4.3 shows the identification results based on three score-level fusion rules in three sub-tables. Each sub-table is also divided into two parts. The upper part shows the recognition rates associated to the facial feature based on individual principal curvature. The lower part displays the recognition rates by fusing two
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Table 4.3: Rank-one identification rate of proposed LPCMP feature descriptor with three score-level fusion rules: the minimum fusion (the upper), the mean fusion (the middle), and the dynamic fusion (the lower).

<table>
<thead>
<tr>
<th>Minimum Fusion</th>
<th>3-ring</th>
<th>5-ring</th>
<th>7-ring</th>
<th>Multi-size Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>85.82%</td>
<td>85.87%</td>
<td>85.70%</td>
<td>88.62%</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>85.87%</td>
<td>85.48%</td>
<td>85.33%</td>
<td>87.93%</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>83.06%</td>
<td>85.39%</td>
<td>84.46%</td>
<td>87.42%</td>
</tr>
<tr>
<td>$\lambda_1 + \lambda_2$</td>
<td>89.56%</td>
<td>90.02%</td>
<td>88.64%</td>
<td>91.05%</td>
</tr>
<tr>
<td>$\lambda_2 + \lambda_3$</td>
<td>88.31%</td>
<td>89.56%</td>
<td>87.48%</td>
<td>90.82%</td>
</tr>
<tr>
<td>$\lambda_3 + \lambda_1$</td>
<td>87.72%</td>
<td>88.72%</td>
<td>86.05%</td>
<td>90.56%</td>
</tr>
<tr>
<td>$\lambda_1 + \lambda_2 + \lambda_3$</td>
<td>90.38%</td>
<td>90.56%</td>
<td>89.84%</td>
<td>92.55%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean Fusion</th>
<th>3-ring</th>
<th>5-ring</th>
<th>7-ring</th>
<th>Multi-scale fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>85.82%</td>
<td>85.87%</td>
<td>85.70%</td>
<td>89.99%</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>85.87%</td>
<td>85.48%</td>
<td>85.33%</td>
<td>89.42%</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>83.06%</td>
<td>85.39%</td>
<td>84.46%</td>
<td>88.91%</td>
</tr>
<tr>
<td>$\lambda_1 + \lambda_2$</td>
<td>90.30%</td>
<td>89.56%</td>
<td>88.41%</td>
<td>91.77%</td>
</tr>
<tr>
<td>$\lambda_2 + \lambda_3$</td>
<td>87.35%</td>
<td>89.31%</td>
<td>89.11%</td>
<td>90.19%</td>
</tr>
<tr>
<td>$\lambda_3 + \lambda_1$</td>
<td>88.62%</td>
<td>89.10%</td>
<td>88.44%</td>
<td>90.19%</td>
</tr>
<tr>
<td>$\lambda_1 + \lambda_2 + \lambda_3$</td>
<td>90.82%</td>
<td>90.46%</td>
<td>90.46%</td>
<td>92.82%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic Fusion</th>
<th>3-ring</th>
<th>5-ring</th>
<th>7-ring</th>
<th>Multi-scale fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>85.82%</td>
<td>85.87%</td>
<td>85.70%</td>
<td>90.30%</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>85.87%</td>
<td>85.48%</td>
<td>85.33%</td>
<td>89.99%</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>83.06%</td>
<td>85.39%</td>
<td>84.46%</td>
<td>89.70%</td>
</tr>
<tr>
<td>$\lambda_1 + \lambda_2$</td>
<td>89.42%</td>
<td>89.56%</td>
<td>88.79%</td>
<td>92.23%</td>
</tr>
<tr>
<td>$\lambda_2 + \lambda_3$</td>
<td>88.62%</td>
<td>90.02%</td>
<td>88.66%</td>
<td>91.05%</td>
</tr>
<tr>
<td>$\lambda_3 + \lambda_1$</td>
<td>88.91%</td>
<td>89.43%</td>
<td>87.93%</td>
<td>90.57%</td>
</tr>
<tr>
<td>$\lambda_1 + \lambda_2 + \lambda_3$</td>
<td>90.72%</td>
<td>90.72%</td>
<td>90.50%</td>
<td>93.16%</td>
</tr>
</tbody>
</table>
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

(1st – 3rd rows) or three (4th row) principal curvature measures with corresponding fusion rule. Furthermore, the fusion result of three region sizes of Borel subsets are also given in last column.

From Table 4.3, when use only one principal curvature measure to identify the subject, the fusion of various region sizes of Borel subset raises 3% to 5% the recognition rate. It indicates that the principal curvature measures estimated in different Borel subsets offer complementary shape information of facial surface. Because the facial surface is a non-homogeneous surface unlike the synthetic geometric surface, it needs a mesh grid with adjustable resolutions to fit different regions on face. But the distribution of vertices and facets in the face scan here is homogeneous. Therefore, the correctness of estimation of principal curvature measures with same Borel subset in different regions varies a bit. When we fuse these principal curvature measures based feature, the better estimation results from the more suitable Borel subset retrieve the lost of other cases.

Furthermore, the recognition rates in every lower part of sub-table are significantly higher than the ones in upper part. This experimental results demonstrate the effectiveness of fusion between different LPCMP features based on principal curvature measures. In three different fusion rules, the fusion of two features increases about 4% of recognition rate, and the fusion of three features increases 5% of recognition rate. It indicates that three principal curvature measures carry meaningful complementary shape information. In addition, even though the experiments have demonstrated that each principal curvature measure indeed carries the meaningful shape information, the combination of $\lambda_{1B}$ and $\lambda_{2B}$ achieves generally higher recognition rate than the combinations involving $\lambda_{3B}$.

Comparing three fusion rules, the dynamic fusion rule enhance relatively more effectively the recognition performance of the proposed 3D face recognition method. Meanwhile, the best rank-one recognition rate as 93.16% is achieved when adopting the dynamic fusion rule. Moreover, the mean fusion rule obtain better performance than minimum fusion rule. Because the latter fusion rule choose the minimum reconstruction residual which makes the identification more rough than the previous one. Based on these experimental results and the analysis above, we decide to perform the verification test on FRGCv2 database with dynamic fusion rule for using three 3D facial feature based on three principal curvature measures.

4.3.3.4 Verification experiments

The verification experiment consists of two modalities: (1) fusion of three features following dynamic rule in single scale of Borel subset; (2) fusion of three features and three scales of Borel subsets. For each subject, the first image was regarded as the gallery and the remaining images are treated as the probe. The True Accept
Chapter 4. 3D Face Recognition with Curvature Faces based on Principal Curvature Measures

Figure 4.6: Receiver operating characteristic (ROC) curves of True Accept Rate (TAR) to False Accept Rate (FAR).

Rate (TAR) and Equal Error Rate (EER) are regarded as the verification criteria. In Fig.4.6, the ROC curve is plotted to show the verification performance of our proposed 3D face recognition method. And Fig.4.7 shows the DET graph in different cases. In first figure, for the case that the feature generated by one principal curvature measure, the verification performance is only tested when the geodesic radius of Borel subset is 5. From two figures, it is obvious to draw similar conclusion as in identification that the fusion of three principal curvature measures can syncretise effectively the shape information of them, and the fusion of multiple scales of Borel subsets can further enhance the performance of LPCMP feature.

In specific, when FAR is 0.001, the TAR are about 88% if we adopt the feature generated by individual principal curvature measure. But the fusions of all three principal curvature measures estimated in three scales of Borel subset increase the TAR to about 92%. Furthermore, the fusion of all principal curvature measures and all scales of Borel subsets achieves 93.33% of verification rate. In this case, the EER is 2.56%.

4.3.3.5 Comparison with the-state-of-the-art

We compare our experimental results with other representative algorithms published performed on FRGC v2.0 database. The comparison is shown in Table 4.4. As we presented in Section 4.1, the maximal curvature and the minimal curvature are used in [Kakadiaris et al. 2007, Wang et al. 2010b, Li et al. 2014a] to describe...
the local geometry of facial surface and build the facial feature descriptor. The result shows that some other methods perform better, while other performs worse. Even the recognition result of our proposed method is not the best, but our experiments have demonstrated that the isolated third principal curvature measure is meaningful for 3D face recognition. It provides complementary shape information for classical principal curvatures. This new geometric attribute pave a new road to the shape analysis based 3D face recognition.
4.4 Conclusion

In this chapter, we introduce a geometry-texture based 3D facial feature, named as Local Principal Curvature Measures Pattern (LPCMP). The principal curvature measures are computed directly on 3D meshed face scan, and then projected to texture image for generating curvature faces. LBP-based encoding pattern update the curvature faces and a local region based feature is extracted on it. SRC is finally adopted to calculate the similarity measurement. In the experiment performing on FRGCv2 database, we evaluate the impact of the region size of Borel subset and the encoding pattern. Towards the previous evaluation, the recognition rates are not sensitive to the change of the scale, but the fusion of multiple scales improves apparently the recognition performance. For the second evaluation, $Q_{2,16}$ is proven to be the most suitable encoding pattern for identification scenario. Moreover, following the dynamic fusion rule, the score-level fusion of three features based on three principal curvature measures enhance effectively the discriminative power of the 3D face recognition system. The comprehensive experimental results demonstrate that the principal curvature measures are able to describe properly the 3D face surface and the corresponding features can characterise the faces. Furthermore, the three principal curvature measures can offer complementary shape information for enhance the performance of 3D face recognition.

However, the drawback of this proposed 3D facial feature descriptor is that the recognition performance depends intensively on the quality of pose normalization. If the 3D face scan can not be normalized to the frontal view, the curvature faces generated from the same subject also varies a lot. Then the large intra-class distance of feature descriptor has significant influence on the 3D face recognition. Considering about this fact, in the next chapter, we will present a normalization free 3D facial feature descriptor, which is also based on principal curvature measures. But this feature descriptor is more robust to head pose change and external occlusion. Furthermore, this feature descriptor can also be utilized to matching the 3D face scan under different resolutions.
Chapter 5

3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Contents

5.1 Introduction ................................................. 83
5.2 Principal Curvature Measures based 3D Face Description and Recognition ................................................. 86
  5.2.1 Framework of PCM-meshSIFT based 3D face recognition . 86
  5.2.2 Principal curvature measures based 3D keypoint detection . 87
  5.2.3 Principal curvature measures based 3D keypoint description . 89
  5.2.4 SRC-based 3D keypoint matching and score level fusion . 91
  5.2.5 Experiments ................................................. 93
5.3 Heterogeneous Cross-Resolution related Experiments .... 105
  5.3.1 Lock3DFace database ........................................ 107
  5.3.2 Signed Distance Function based facial surface optimization . 108
  5.3.3 Landmarks location with modified Active Shape Models . 109
  5.3.4 Experiments ................................................. 112
5.4 Conclusion .................................................. 116

5.1 Introduction

The analogue methods based on 3D shape analysis have also been widely addressed in the literature of 3D face recognition. Facial surface quantities, such as points [Bronstein et al. 2003], curves [Samir et al. 2006], stripes [Beretti et al. 2010], regions [Chang et al. 2005, Chang et al. 2006], normals [Tang et al. 2013, Li et al. 2014a], curvatures [Tang et al. 2015, Li et al. 2015],
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Geodesic distance [Bronstein et al. 2004, Li & Zhang 2007, Li et al. 2009] have been commonly used to represent and characterize the geometry structures of 3D faces. Among these quantities, surface curvatures, including Gauss curvature, mean curvature, principal curvatures, as well as their variations, namely shape index, are the most widely used ones for 3D face recognition [Gordon 1992, Tanaka et al. 1998, Wu et al. 2004, Li & Zhang 2007, Li et al. 2009, Li et al. 2015, Smeets et al. 2013, Conde et al. 2006, Chua & Jarvis 1997, Li et al. 2014a, Tang et al. 2015]. Thus, the representation and the characterization of 3D face surface are the most important parts for an effective 3D face recognition system. And curvatures play a fundamental effect in the literature of 3D face recognition.

As aforementioned, 3D facial surface is generally represented in point clouds, depth images, and triangle meshes. These discrete surfaces set hindrance to estimate curvatures with classical methods defined in smooth surface. To handle with this problem, we proposed to use principal curvature measures to describe the shape of facial surface. As introduced in Chapter 3, the calculation of principal curvatures is generalized to triangle meshed model by estimating their measure form (i.e. principal curvature measures) in a neighbourhood region (i.e. the Borel subset) associating to a vertex on facial surface. The advantages of this method include that the principal curvature measures represent directly the shape characteristics of 3D meshed face scan, and the changeable estimation region scale can fit the meshed model in different resolutions. In Chapter 4, we designed a principal curvature measures related feature descriptor, and evaluate its discriminative power in 3D face recognition. However, the normalization and alignment are obligatory processing step in that approach, which makes it weakly robust to pose change, external occlusion or partial face missing case.

In this chapter, we propose a registration-free local feature descriptor combining the principal curvature measures and principal curvature vectors. Inspired to SIFT-based facial descriptor [Lowe 2004], which is an outstanding algorithm in 2D face recognition, we present the Principal Curvature Measures based SIFT feature descriptor for 3D meshed face scan (PCM-meshSIFT). This method makes full use of the direction invariance property of principal curvature vectors, and the shape-based description ability of principal curvature measures, by assigning a canonical direction for a vertex and building a Histogram of Curvature (HOC) feature to represent its shape-based characteristics. The comprehensive experiments are performed on FRGCv2 database and Bosphorus database involving pose changes with various rotation angles and along different rotation axis, and external occlusion leading to partial face missing.

Furthermore, the general evaluation of face recognition algorithm utilises the high-resolution face samples which cost too much in data acquisition and matching computation. As aforementioned, for achieving the high-resolution facial model,
the complex data acquisition device is required to generate the structure light or the laser ray, and to capture the shape information by locating the coordinates of massive points on face. Theoretically, more recorded vertices and facets existing in face data can increase the representation precision of facial shape. But the following problem is thus more calculation complexity is required in preprocessing, feature extraction and surface matching. However, the popularisation of customer-level capture device, e.g. Kinect series of camera, offers more opportunities to collect the 3D face models in low-resolution. This data collection pattern is more efficient and the processing on the captured data cost much less. However, the face scan in low-resolution contains sparser vertices and less facets, which limits the richness of shape information represented in facial surface. A possible solution is combining the high-resolution and the low-resolution face sample to achieve enough shape information for recognition and raise the processing efficiency concurrently. In this chapter, we propose to register the high-resolution samples as gallery set, and regard the low-resolution samples as probe set. This recognition pattern leaves the time-consuming part with high-resolution samples into the off-line operation part, and the on-line operation part with low-resolution samples costs less time. Meanwhile, PCM-meshSIFT feature descriptor is adopted to characterize the facial surface and the Sparse Representation based Classifier is used to matching the keypoints. The merit of this method is that the adjustable region scale of Borel subset for estimating principal curvature measures can fit this heterogeneous cross-resolution recognition case. The recognition experiment about the heterogeneous cross-resolution samples is performed on Lock3DFace database, which contains large scale of both high-resolution and low-resolutions 3D face samples.

In the following parts, we firstly introduce the PCM-meshSIFT based 3D facial feature descriptor, including 3D keypoint detection, 3D keypoint description, and 3D keypoint matching. The homogeneous face recognition experiments are then presented, including the ones performed on FRGCv2 database and Bosphorus database. They evaluate the robustness ability and generalization ability of our proposed feature to pose and occlusion variations. Then, we apply the PCM-meshSIFT feature on heterogeneous face recognition between face samples in different resolutions. The specific preprocessing technique and keypoint location method are presented briefly in that part. Finally, we present the experiments based on the heterogeneous cross-resolution 3D face recognition on Lock3DFace database.
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

5.2 Principal Curvature Measures based 3D Face Description and Recognition

5.2.1 Framework of PCM-meshSIFT based 3D face recognition

To show the effectiveness of principal curvature measures for 3D face description and recognition, we follow the popular 3D surface matching framework [Guo et al. 2016, Guo et al. 2014, Li et al. 2015, Smeets et al. 2013]. As shown in Fig.5.1, the description and recognition framework includes three basic modules: 3D keypoint detection, 3D keypoint description and 3D keypoint matching.

For 3D facial keypoint detection, we propose to build three principal curvature measures based surface Gaussian scale space to ensure the scale invariant property of the keypoint description. 3D keypoints are detected by finding the extrema in Difference of Curvature (DOC), inspired to Difference of Gaussian (DOG), across scales and one-ring neighbors. Since we use three principal curvature measures, it expected that our detection method can locate more stable and meaningful keypoints.

For 3D facial keypoint description, to comprehensively and robustly describe the local shape around a detected 3D keypoint, we compute the statistical histograms of all the three principal curvature measures within a geodesic neighbourhood of the detected keypoint. To make the local shape descriptor robust to head pose variations, it needs to build a local coordinate system for each keypoint. In this approach, we propose to use the statistical property of the three principal curvature vectors to find three local canonical directions for each principal curvature measure. This kind of 3D keypoint description is similar in spirit to the 2D SIFT [Lowe 2004], 2.5D SIFT [Lo & Siebert 2009] and meshSIFT related algorithms [Maes et al. 2010, Smeets et al. 2013], but the core geometric attributes \((\lambda_1, \lambda_2, \lambda_3)\) and \((e_1, e_2, e_3)\) are more suitable and reasonable for triangle mesh based 3D face descriptor.

For 3D facial keypoint matching, we use multi-task sparse representation [Wright et al. 2009]. This method firstly figures out the sparsest representation of each testing descriptor from the complete dictionary set of all the descriptors in gallery. Then the average reconstruction errors are designed as the similarity measurement between probe scan and gallery scan. So the identity of testing subject is labelled as the gallery subject who has the minimal reconstruction error. The details of each step are introduced in the following subsections.
5.2.2 Principal curvature measures based 3D keypoint detection

In contrast to the LBP-like facial feature presented in Chapter 4, the PCM-meshSIFT based facial feature is extracted on the vertices of 3D meshed face scans. For increasing the efficiency of processing, the best choice is to locate the keypoints on the facial meshed surface, and then extract the feature on them.

The Principal curvature measures based 3D keypoint detection starts by performing a series of surface Gaussian convolutions with different variances (i.e. different scales) over a given 3D face scan in triangle mesh. Given a vertex \( v_i \) of a face model \( F \), the Gaussian convolution smooths the geometry structure of \( F \) over \( v_i \)'s adjacent region and leads to the updated vertex of \( v_i \) as

\[
v_i^{\sigma_s} = \frac{\sum_{v_j \in N(v_i, 1)} g_{\sigma_s}(v_i, v_j) \cdot v_j}{\sum_{v_j \in N(v_i, 1)} g_{\sigma_s}(v_i, v_j)}
\]

where \( \sigma_s \) denotes the Gaussian convolution variance, \( N(v_i, 1) \) denotes the set of vertices within 1-ring neighbors of \( v_i \), and the Gaussian kernel \( g_{\sigma_s} \) is defined as

\[
g_{\sigma_s}(v_i, v_j) = \exp(-\frac{\|v_i - v_j\|^2}{2\sigma_s^2})
\]

Once the scale space of facial surface is constructed, we estimate the principal curvature measures over the face scan in each scale space, including the original one. Then we compute the Differences of Curvature \( \delta_{\lambda_c} \) related to these principal
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Figure 5.2: The distribution of three principal curvature measures $\lambda_{1B}$, $\lambda_{2B}$ and $\lambda_{3B}$ estimated on 3D meshed face. The second row displays the histograms of the value of principal curvature measures on the face.

curvature measures. The Differences of Curvature is defined as:

$$\delta \lambda_i (v_{s}) = \lambda_i (B_{v_s}) - \lambda_i (B_{v_{s-1}}), i = 1, 2, 3$$

where $v_{s}$ and $v_{s-1}$ denotes the updated coordinates in different scale spaces. 3D keypoints are detected by finding the extreme of $\delta$ among the one-ring vertices $v_j \in N(v_i, 1)$ in $\sigma_s$ and its two adjacent scales $\sigma_{s-1}$ and $\sigma_{s+1}$. If a vertex is located as keypoint, its detection scale is $\sigma_s$. Figure 5.2 illustrates the distribution of three principal curvature measures estimated on the original face scan and their corresponding global histogram. We particularly point out that, in the third histogram plot, the value of the third principal curvature measure centralized around 0, but not equal to 0. Recalling the presentation of principal curvature measures in Chapter 3, the third principal curvature measure is 0 if the Borel subset is reduced to a point. The histogram of this principal curvature measure also verifies the correctness of the estimation. Meanwhile, the map of this principal curvature measure highlight some meaningful parts on face, e.g. the eye corners, the eye orbit, the nose edges, and the mouth. It demonstrates that the coherence of the generalization and the existence of the special shape information carried by this supplementary principal curvature measure.

Figure 5.3 shows the detected 3D facial keypoints by using the three principal curvature measures, respectively. For each principal curvature measure, around 476 keypoints, which is 1.36% of the quantity of vertices in whole face, are detected for a given 3D face scan. It also means the calculation efficiency is 73 times theoreti-
3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Figure 5.3: Illustration of detected 3D keypoints using principal curvature measures $\lambda_1$, $\lambda_2$, and $\lambda_3$ (from left to right), respectively.

cally faster than using all facial vertices to describe the 3D face model. Moreover, the detected 3D keypoints distribute in the typical face parts and some rigid transformation areas. It ensures the identification accuracy even there exists head pose variations.

5.2.3 Principal curvature measures based 3D keypoint description

5.2.3.1 Canonical directions assignment

In 3D face recognition, the pose variations have intensive impact on the recognition performance of a feature descriptor by increasing significantly the intra-class distance. There are two possible solutions to handle with this problem: (1) Adding a supplementary pose alignment step in the registration eliminates the interference of pose changes; (2) Designing a rotation invariant local shape descriptor avoids the dissimilarity of descriptors extracted from the facial scans in different poses. In order to guaranteeing a higher processing efficiency, we apply to determine the canonical direction based on principal curvature vectors of each keypoint for achieving a rotation-invariance descriptor remaining robust to the pose variations.

Given a meshed 3D facial surface $T$, for each keypoint $v_k$, the vertices $v_j$ situating in its neighbourhood are selected to become the support for assigning the canonical direction. The neighbourhood of $v_k$ is defined as:

$$\mathcal{N}(v_k) = \{v_j \in T | d_g(v_k, v_j) < R\}$$

(5.4)

where $d_g(v_k, v_j)$ denotes the geodesic distance between $v_k$ and $v_j$, and $R$ is set as $9\sigma_s$.

Then, a plane $\Pi_{v_k}$ through $v_k$ orthogonal to $e_3(B_{v_k})$ is built as a common plane for the canonical directions based on three principal curvature vectors. We choose
where $w(1 \text{ bin per } 1^\circ)$. The weight $v$ mesh approximates to a smooth surface. Due to the principal curvature vectors of $v_k$ are the eigenvectors of the real symmetric matrix $M_B^T$, three principal curvature vectors are orthogonal. A local coordinate system is hence built as $(e_1(B_{v_k}), e_2(B_{v_k}), e_3(B_{v_k}))$.

All the selected $v_j$ and their principal curvature vector $e_i(B_{v_j})$ are transformed into this local coordinate system and further projected to $\Pi_{v_k}$, and the projections are denoted as $\tilde{v}_j$ and $\tilde{e}_i(B_{\tilde{v}_j})$, respectively. The magnitude of $\tilde{e}_i(B_{\tilde{v}_j})$, and the angle between $\tilde{e}_i(B_{\tilde{v}_j})$ and $e_1(B_{v_k})$ are defined as:

$$
mag(\tilde{e}_i(B_{\tilde{v}_j})) = \sqrt{\tilde{e}_i^x(B_{\tilde{v}_j})^2 + \tilde{e}_i^y(B_{\tilde{v}_j})^2}$$

$$
\theta(\tilde{e}_i(B_{\tilde{v}_j})) = \arctan(\tilde{e}_i^y(B_{\tilde{v}_j})/\tilde{e}_i^x(B_{\tilde{v}_j})),$$

where $\tilde{e}_i^x(B_{\tilde{v}_j}) = e_1(B_{v_k}) \cdot \tilde{e}_i(B_{\tilde{v}_j})$ and $\tilde{e}_i^y(B_{\tilde{v}_j}) = e_2(B_{v_k}) \cdot \tilde{e}_i(B_{\tilde{v}_j})$.

Finally, we build a weighted histogram of $\theta(\tilde{e}_i(B_{\tilde{v}_j}))$, and divide it into 360 bins (1 bin per $1^\circ$). The weight $w(\tilde{v}_j)$ is defined as:

$$
w(\tilde{v}_j) = \text{mag}(\tilde{e}_i(B_{\tilde{v}_j})) \cdot g(d_e(\tilde{v}_j, v_k)),$$

where $g$ denotes the Gaussian kernel with the standard deviation set as half of $R$, and $d_e(\tilde{v}_j, v_k)$ denotes the Euclidean distance between $\tilde{v}_j$ and $v_k$. The $e_i(B_{v_k})$ based canonical direction $d_i(v_k)$ is assigned as the peak of the weighted direction histogram.

In the following keypoint description step, we will extract 9 histograms in 9 neighbouring circular regions around the keypoint and generate the local descriptor by jointing them together. These 9 regions also situate in the plane $\Pi_{v_k}$. The canonical direction will be used to define the locations of these regions for guaranteeing the direction invariance of the local descriptor. It is the reason that the canonical direction assigned here lies in the projected plane $\Pi_{v_k}$ rather than in $\mathbb{R}^3$.

### 5.2.3.2 3D keypoint descriptor representation

As shown in Fig. 5.5, descriptor configuration is performed on the plane $\Pi_{v_k}$ of the new local coordinate systems. Inspired by the 2D daisy descriptor [Tola et al. 2010], 9 overlapping circular regions with a radius of $r = 3.75\sigma$ are assigned centering at the keypoint $v_k$ and its 8 neighbouring points, respectively (see Fig. 5.5). This kind of quasidaisy radial flower pattern of overlapping circles simulates the functioning of human complex cells in the visual cortex [Hubel & Wiesel 1962b]. Therefore, it tends to be robust to small transformations, e.g. spatial shifting, non-rigid deformations. The 8 neighbouring points are localized by performing uniform sampling over
a circle centered at $v_k$ with a radius of $R = 4.5\sqrt{2}\sigma_s$, starting from one canonical direction $d_i(Bv_k)$ corresponding to one principal curvature measure $\lambda_{i_B}$.

In each circular region $r_i (i = 1, 2, \cdots, 9)$, we construct the local weighted histograms of different principal curvature measures: $\lambda_{1_B}$, $\lambda_{2_B}$, and $\lambda_{3_B}$. Their corresponding histograms are referred to as the histogram of the first principal curvature measure ($hoc_1^i$), histogram of the second principal curvature measure ($hoc_2^i$), and histogram of the third principal curvature measure ($hoc_3^i$), respectively.

The values of principal curvature measures are quantized equally to 8 bins, and weighted by a Gaussian kernel $g_{\sigma_s}(r_i, r_j)$, where the standard deviation is set as the Euclidian distance between the current point and the center point of the circle region. The final histograms at keypoint $v_k$ are constructed by concatenating $hoc_1^i$, $hoc_2^i$, and $hoc_3^i$ in a clockwise direction, represented as:

$$
\begin{align*}
HOC_1 &= (hoc_1^i, hoc_2^i, \cdots, hoc_9^i), \\
HOC_2 &= (hoc_1^i, hoc_2^i, \cdots, hoc_9^i), \\
HOC_3 &= (hoc_3^i, hoc_3^i, \cdots, hoc_3^i).
\end{align*}
$$

The above sub-histograms (e.g., $hoc_1^i$) and histograms (e.g., $HOC^i$) are all normalized to unit vectors to eliminate the influence of non-uniform mesh sampling. This generates three 3D keypoint descriptors with the same dimension of 72. We will show that these three local descriptors contain strong complementarity information in descriptiveness.

### 5.2.4 SRC-based 3D keypoint matching and score level fusion

Based on the principal of Sparse Representation based Classifier [Wright et al. 2009], Li et al. [Li et al. 2015] proposed the SRC-based 3D keypoint matching method, which has been successfully used for 3D face recognition. The
first step to perform the SRC-based matcher is to build a dictionary, including all the 3D keypoint descriptors of the gallery subjects. Given a gallery set of $N$ subjects, each subject has a single 3D face scan. Assume that $n_i$ 3D keypoints are detected for $i$-th subject. Let the corresponding $n_i$ 3D keypoint descriptors be the following sub-dictionary:

$$D_i = [d_{i,n_1}, d_{i,n_2}, \cdots, d_{i,n_i}] \in \mathbb{R}^{m \times n_i}, \quad (5.8)$$

where $m$ is the descriptor dimension. Then, the dictionary for all the $N$ subjects of the gallery can be built by simply concatenating all the sub-dictionaries as:

$$D = [D_1, D_2, \cdots, D_N] \in \mathbb{R}^{m \times K}, \quad (5.9)$$

where $K = n_1 + n_2 + \cdots + n_N$ represents the total number of keypoint descriptors in the gallery.

Once the gallery dictionary has been constructed, one needs to solve the following multi-task sparse representation problem:

$$\arg \min_{x_i} \|y_i - Dx_i\|_2 \quad s.t. \quad \|x_i\|_0 \leq L, \quad i = 1, 2, \cdots, n \quad (5.10)$$

where $y_i$ denotes $i$-th 3D keypoint descriptor of a probe face scan, $x_i$ is the sparse coefficient, and $\| \cdot \|_0$ denotes the $l_0$ norm of a vector, defined as the number of non-zero elements of the vector, $L$ is the sparsity parameter, which controls the sparsity of the solution.
Finally, the identity of the probe face scan can be determined by the following average accumulative sparse reconstruction error, as the similarity measurement, for all the keypoints of a given probe face scan:

\[
\text{identity}(Y) = \arg \min_j \frac{1}{n} \sum_{i=1}^{n} \|y_i - D\delta_j(\hat{x}_i)\|_2^2,
\]  

(5.11)

where \(\delta_j(\cdot)\) is a characteristic function which selects only the coefficients associated with the \(j\)-th subject.

As mentioned before, three principal curvature measures and vectors are used to form three feature descriptors. We combine their similarity measurements with score-level fusion rules to combine all their contributions for final decision step.

### 5.2.5 Experiments

We evaluate the effectiveness of the proposed approach on both the FRGC v2.0 database [Phillips et al. 2005] and the Bosphorus database [Savran et al. 2008]. The FRGCv2 database is the popular benchmark in 3D face recognition, which is suitable to check the discriminative power and generalization skill of the proposed approach. The Bosphorus database is the largest public 3D face database with various challenges in uncontrolled condition of 3D face recognition, including expression variations, pose changes and external occlusions.

In this section, we will present these two databases briefly and the related experiment settings. The experiment results related to two databases will be presented respectively to show the recognition performance of PCM-meshSIFT feature descriptor under various experimental conditions.

#### 5.2.5.1 Databases

In the following experiments, we test the recognition performance of the proposed approach on the FRGCv2 database and the partial Bosphorus database. About the introduction of the FRGCv2 database, please refer to Section 4.3.1. Bosphorus database is quite unique for its broad set of expression types, systematic variation of poses and different external occlusions. It depicts many variations challenges which may typically occur in the real-life application under uncontrolled environment. Specifically, Bosphorus database contains totally 4666 3D face scans of 105 subjects, involving 44 women and 61 men. For each subject, there are around 34 expressions in different intensities, 13 poses with head rotation in single or multiple rotational degrees of freedoms, and 4 kinds of external occlusions. In the following performance evaluation, we will use the facial samples with different head poses and all types of occlusions, but without expressions.
Figure 5.6: Samples of head pose variations in Bosphorus database. (a) frontal; (b) Yaw +10°; (c) Yaw +20°; (d) Yaw +30°; (e) Yaw +45°; (f) Yaw -45°; (g) Yaw +90°; (h) Yaw -90°; (i) Pitch upward rotation (PR); (j) Pitch downward rotation (PR); (k) Cross rotation (CR): yaw rotation and approximately pitch rotation +20°; (l) Cross rotation (CR): yaw rotation and approximately pitch rotation -20°.

(i) Head pose: As shown in Fig.5.6, this subset contains 7 Yaw Rotations (YR): +10°, +20°, +30°, +45°, -45°, +90°, -90°; 4 Pitch Rotations (PR): strong upwards/downwards, slight upwards/downwards; 2 Cross Rotations (CR): yaw rotation +45° and approximately pitch rotation ±20°. Each subject possesses one samples with each rotation pattern of the 7 patterns mentioned.

(ii) Occlusions: As shown in Fig.5.7, for each subject, 4 types of samples with external occlusion, including the occlusion by glasses (Glasses), occlusion by hair (Hair), occlusion on eye part by hand (HandE), and Occlusion on mouth by hand (HandM). Particularly, the occlusion occurred by hair including the head hair and the facial hair like beard and moustache in male samples. In total, there are 105 samples of each kind of occlusion, except the occlusion by hair which are only 67 samples.

The 3D face scans in Bosphorus database were captured by 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 contains approximately 35K vertices.
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Figure 5.7: Samples of external occlusion in Bosphorus database. (a) Occlusion by glasses; (b) Occlusion by hair; (c) Occlusion on eye part by hand; (d) Occlusion on mouth by hand.

5.2.5.2 Experimental settings

For the experiments on FRGCv2 database, each subject has 2~22 samples with different facial expressions. We take the earliest face scan of each person to group the gallery set (466 face scans) and the rest forms the probe gallery (3,541 face scans).

(i) Experimental configuration on the FRGC v2.0 database

On the first subset, three score-level fusion methods (mean rule, product rule and minimum rule) of three individual feature descriptors (HOC$_1$, HOC$_2$, and HOC$_3$) are tested. This fusion operation is to verify the practicality and informative complementarity of three principal curvature measures. Meanwhile, according to the region size of Borel subset used to estimate principal curvature measures, there are three experiment branches corresponding to three region sizes. The same score-level fusion method is also performed to combine the feature descriptor associated to different region sizes.

In summary, the 3D face recognition experiments are carried out from five viewpoints: (a) Identification with one principal curvature measure HOC$_i$, (i = 1, 2, 3) as baseline; (b) Identification with fusion of HOC$_i$, HOC$_j$, (i, j = 1, 2, 3, i ≠ j); (c) identification with fusion of HOC$_1$, HOC$_2$ and HOC$_3$; (d) In the previous experiment scenarios, the principal curvature measures are estimated in three region size of Borel subset, i.e. 3-ring, 5-ring and 7-ring, which are the geodesic radii of Borel subset respectively; (e) Three kinds of score-level fusion rules are examined in all the fusion operations aforementioned for both identification and verification.

(ii) Experimental description on the Bosphorus database

For the experiments on Bosphorus database, we take 105 samples with neutral expression (1 sample for each subject) to construct the gallery set at first. The remaining neutral expression samples are regarded as Neutral (N) expression probe set. All samples with pose changes (P) and external occlusion (O) are also regarded
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Table 5.1: The distribution of 3D face scans over the various probe subsets in the Bosphorus database.

<table>
<thead>
<tr>
<th>Expression</th>
<th>Occlusion</th>
<th>Pose Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutral</td>
<td>HandE</td>
</tr>
<tr>
<td>Neutral</td>
<td>194</td>
<td>105</td>
</tr>
<tr>
<td>Pose Changes</td>
<td>YR 10°</td>
<td>YR 20°</td>
</tr>
<tr>
<td>YR 10°</td>
<td>105</td>
<td>105</td>
</tr>
</tbody>
</table>

as probe set with uncontrolled variations. The details of various probe subsets are listed in Tab.5.1.

In the identification scenario on Bosphorus database, we also evaluate the performance of three individual feature descriptor and the practical function of their score-level fusion. Only the mean fusion rule is utilised to perform the fusion operation. Moreover, because the effectiveness of multi-scale fusion has been confirmed in previous experiment, we adopt this fusion strategy directly to simplify the evaluation process. Similarly to the experiments on FRGCv2 DB, the multi-scale based fusion is also performed for fusing the descriptor based on three region scales (i.e. the geodesic radius 3, 5 and 7).

5.2.5.3 3D face recognition on FRGCv2 database

In this part, we present the experimental results of the principal curvature measures based 3D face description and recognition framework carrying on FRGCv2 database. Several experimental scenarios are performed to evaluate the effectiveness of single proposed feature descriptor in 3D face recognition, the impact of region scale of Borel subset, the discriminative power enhanced by fusing the features and the region scales. Furthermore, our experimental results are compared to the state-of-the-arts.

(1) Baseline evaluation with three feature descriptors individually

We firstly evaluate the baseline identification rate of three feature descriptors based on three individual principal curvature measures and vectors respectively. Tab.5.2 lists the rank-1 recognition rates of three feature descriptors associated to three region scales of Borel subset (2nd to 4th columns), and the multi-scale based fusion following three score-level fusion rules (5th to 7th columns).

As shown in Tab.5.2, three 3D facial feature descriptors can verify the subject’s identity individually. Without the multi-scale fusion strategy, the recognition rates of three feature descriptor are above 82%. This results demonstrate that all three principal curvature measures and their corresponding 3D facial features are able to
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Table 5.2: The baseline identification rates of 3D FR with individual principal curvature measure based feature descriptor on FRGCv2 database

<table>
<thead>
<tr>
<th>Baseline</th>
<th>3-ring</th>
<th>5-ring</th>
<th>7-ring</th>
<th>Multi-Scale Fusion</th>
<th>Mean</th>
<th>Product</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOC₁</td>
<td>91.34%</td>
<td>92.62%</td>
<td>90.07%</td>
<td>92.92%</td>
<td>91.42%</td>
<td>91.26%</td>
<td></td>
</tr>
<tr>
<td>HOC₂</td>
<td>83.20%</td>
<td>85.49%</td>
<td>85.71%</td>
<td>87.46%</td>
<td>86.33%</td>
<td>85.90%</td>
<td></td>
</tr>
<tr>
<td>HOC₃</td>
<td>84.48%</td>
<td>85.24%</td>
<td>82.44%</td>
<td>86.82%</td>
<td>85.47%</td>
<td>84.69%</td>
<td></td>
</tr>
</tbody>
</table>

classify the local shape of facial surface. We also observe that the feature descriptor based on the first principal curvature measure (i.e. HOC₁) performs much better than other two feature descriptors (i.e. HOC₂ and HOC₃), whose recognition rate is approximately 91.34%, while other two feature descriptors achieve 84.80% and 84.05%.

Furthermore, in this approach, the region size of Borel subset has marginally impact on the recognition rates. The principal curvature measures estimated in the 5-ring Borel subset has more discriminative power. With this region scale, the accuracy of HOC₁ and HOC₃ based feature are higher than the ones with other two scales of 1% to 3%. But the exception appears in the case when using HOC₂ based feature with the 7-ring Borel subset, which is a bit higher than the one of 5-ring. Therefore, it is hardly to predicate determinately that the 5-ring of Borel subset is the best scale for 3D face recognition on FRGCv2 database. However, the fusion of multi-scale Borel subsets improves the identification ability of three 3D facial feature descriptor. The increment intensity of recognition rate is more observable with the features based on HOC₂ and HOC₃. Especially, the fusion of multi-scale Borel subsets following mean fusion rule raises the recognition rates of three individual feature descriptor as 1.58%, 2.66% and 2.77%.

(2) Evaluation of the fusion of three feature descriptors

According to the dissimilarity of distribution of three principal curvature measures on face, they highlight different local region on face and their related feature descriptors should also possess the unique response to different local facial shapes. In other words, if we combine three feature descriptor, the complementary shape information quantized by three principal curvature measures is expected to ameliorate the recognition performance synchronously. Based on this speculation, we perform the experiments by fusing three feature descriptors and list the rank-one recognition rates in Table 5.3. There are three sub-tables related respectively to three score-level fusion rule (i.e. mean rule, product rule and minimum rule). Following each score-level fusion rule, we also design four fusion pattern for feature descriptors, including three one-to-one combinations (i.e. (HOC₁ + HOC₂), (HOC₂ + HOC₃), (HOC₃ + HOC₁)) and a global combination (i.e. (HOC₁ + HOC₂ + HOC₃)).

97
similar to Tab.5.2, all the feature descriptors are constructed in three region scale of Borel subsets, and they also are fused following the corresponding fusion rule.

Comparing to the listed results in Tab.5.2, the score-level fusion method increases intensively the rank-one recognition rates achieved by individual feature descriptor. Let us focus on the mean fusion rule with 3-ring, the combination of \((HOC_1 + HOC_2)\) achieves 96.18\% the recognition rate, the combination of \((HOC_2 + HOC_3)\) obtains 92.88\% the recognition rate, and the combination of \((HOC_3 + HOC_1)\) achieves the recognition rate of 96.18\%. These recognition accuracy is obviously higher than the ones (91.34\%, 83.20\% and 84.48\%) achieved by using the feature descriptor individually. If we fuse all three feature descriptors, the rank-one recognition rate raise to 97.46\%. The similar increments of recognition rate can also be observed in the remaining scenarios with radius 3 (resp. 7) and product rule (resp. minimum rule). Moreover, as shown in Tab.5.2, the discriminative power of \(HOC_2\) and \(HOC_3\) related feature is weaker than \(HOC_1\) related feature, but their combination can also achieve promising recognition performance. These fusion based experimental results demonstrate our speculation that there exists the complementary shape information carried by three principal curvature measures, and the PCM-meshSIFT feature descriptors can quantize these complementary shape information for 3D face recognition.

Similarly to the baseline evaluation, the fusion of multi-scale of Borel subset can also raise the recognition rates in various intensity. Due to the structure of facial surface is non-homogeneous surface, different local regions need to be represented by the set of facets in various density. But in our experiment, the distribution of vertices and facets in the face model is homogeneous. Therefore, the correctness of estimation of principal curvature measures with same Borel subset in different regions varies a bit. When we fuse these principal curvature measures based feature, the better estimation results from the more suitable Borel subset retrieve the lost of other cases.

Furthermore, three kinds of fusion rules are evaluated to fuse \(HOC_1\), \(HOC_2\), and \(HOC_3\) based keypoint descriptors. From the results, we can observe that all three kinds of fusion rules can effectively enhance the performance of our proposed method. Among them, the mean rule perform slightly more stable than the other two fusion rules. Meanwhile, with the mean fusion rule, we achieve the highest recognition rate as 97.96\%.

We also plot the cumulative match characteristic (CMC) curve of individual descriptors \((HOC_1, HOC_2, \text{and } HOC_3)\) and their score-level fusion with mean fusion rule \((HOC_1 + HOC_2 + HOC_3)\) in Fig 5.8. The combination of three principal curvature measures can achieve outstanding identification results, and increase visibly the discriminative power. This experimental results prove evidently the complementary informative of three principal curvature measures based facial description,
Table 5.3: Recognition performance evaluation of different fusion combinations of PCM-meshSIFT feature descriptor on the FRGCv2 database. Three score-level fusion rules are listed: Mean rule (upper sub-table), Product Rule (middle sub-table), and Minimum rule (lower sub-table).

<table>
<thead>
<tr>
<th>Fusion Rule</th>
<th>3-ring</th>
<th>5-ring</th>
<th>7-ring</th>
<th>Multi-Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Rule</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOC1, HOC2</td>
<td>96.18%</td>
<td>95.42%</td>
<td>92.11%</td>
<td>97.49%</td>
</tr>
<tr>
<td>HOC2, HOC3</td>
<td>92.88%</td>
<td>93.64%</td>
<td>90.33%</td>
<td>93.59%</td>
</tr>
<tr>
<td>HOC3, HOC1</td>
<td>96.18%</td>
<td>97.20%</td>
<td>91.60%</td>
<td>97.66%</td>
</tr>
<tr>
<td>HOC1, HOC2, HOC3</td>
<td>97.46%</td>
<td>97.96%</td>
<td>93.38%</td>
<td>97.96%</td>
</tr>
<tr>
<td>Product Rule</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOC1, HOC2</td>
<td>96.48%</td>
<td>95.42%</td>
<td>92.36%</td>
<td>97.04%</td>
</tr>
<tr>
<td>HOC2, HOC3</td>
<td>91.86%</td>
<td>91.35%</td>
<td>90.84%</td>
<td>92.52%</td>
</tr>
<tr>
<td>HOC3, HOC1</td>
<td>96.44%</td>
<td>95.92%</td>
<td>92.31%</td>
<td>96.87%</td>
</tr>
<tr>
<td>HOC1, HOC2, HOC3</td>
<td>96.95%</td>
<td>97.20%</td>
<td>92.62%</td>
<td>97.46%</td>
</tr>
<tr>
<td>Minimum Rule</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOC1, HOC2</td>
<td>95.17%</td>
<td>94.90%</td>
<td>92.11%</td>
<td>96.42%</td>
</tr>
<tr>
<td>HOC2, HOC3</td>
<td>93.13%</td>
<td>91.09%</td>
<td>89.31%</td>
<td>92.88%</td>
</tr>
<tr>
<td>HOC3, HOC1</td>
<td>95.16%</td>
<td>95.17%</td>
<td>92.36%</td>
<td>96.18%</td>
</tr>
<tr>
<td>HOC1, HOC2, HOC3</td>
<td>96.69%</td>
<td>97.46%</td>
<td>95.67%</td>
<td>97.87%</td>
</tr>
</tbody>
</table>

and the feasibility of our proposed 3D face identification framework.

The fusion of three principal curvature measures based facial features also enhance the verification performance of the 3D face recognition method. As displayed in Fig.5.9, the number of the successful verified samples increases obviously when we combine the three facial features. The highest verification rate is 95.28% when False Accept Rate is 0.001, which is significant higher than the 90.03% verification rate if we only use the first principal curvature measure based feature. With three facial features, the EER is 2.12%.

(4) Comparison with the state-of-the-art approaches

In Table 5.4, we compare our identification and verification results with other state-of-the-art 3D face recognition approaches on the whole FRGCv2 database. Note that for 2.5D range image based approaches (upper part), Spreeuwers’s [Spreeuwers 2015] approach achieved the highest rank-one recognition rate of 99.4% and the highest verification rate of 99.3%. For 3D face triangle mesh based approaches, the 98.37% rank-one recognition rate and the 95.28% verification rate of our method are the competitive ones among the analogue 3D face recognition approaches. Furthermore, all these methods need sophisticated and time-consuming 3D face registration algorithm. Our framework is totally registration-free. When focusing on the curvature-based 3D face recognition approaches, e.g. the ones in [Li et al. 2015, Smeets et al. 2013, Tang et al. 2015], our proposed principal curvature measures based 3D face recognition framework also achieves the best perfor-
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Figure 5.8: Cumulative Match Characteristic of HOC1, HOC2, HOC3 and their fusions from rank 1 to 10.

Figure 5.9: ROC curve with one individual principal curvature measures based feature and the fusion of three features.
Table 5.4: Performance comparison on the FRGC v2.0 database.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Year</th>
<th>Data format</th>
<th>Rank-1 TAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Chang et al. 2006]</td>
<td>2006</td>
<td>Range</td>
<td>91.90%</td>
</tr>
<tr>
<td>[Mian et al. 2007]</td>
<td>2007</td>
<td>Range</td>
<td>95.73% 98.31%</td>
</tr>
<tr>
<td>[Kakadiaris et al. 2007]</td>
<td>2007</td>
<td>Range</td>
<td>97.00% 97.30%</td>
</tr>
<tr>
<td>[Wang et al. 2010b]</td>
<td>2010</td>
<td>Range</td>
<td>98.30% 97.97%</td>
</tr>
<tr>
<td>[Spreeuwers 2011]</td>
<td>2011</td>
<td>Range</td>
<td>99.00% 94.60%</td>
</tr>
<tr>
<td>[Huang et al. 2011b]</td>
<td>2011</td>
<td>Range</td>
<td>97.20% 98.40%</td>
</tr>
<tr>
<td>[Spreeuwers 2015]</td>
<td>2015</td>
<td>Range</td>
<td>99.40% 99.30%</td>
</tr>
<tr>
<td>[Ballihi et al. 2012]</td>
<td>2012</td>
<td>Mesh</td>
<td>98.02%</td>
</tr>
<tr>
<td>[Li et al. 2015]</td>
<td>2013</td>
<td>Mesh</td>
<td>96.30%</td>
</tr>
<tr>
<td>[Smeets et al. 2013]</td>
<td>2013</td>
<td>Mesh</td>
<td>89.60% 79.00%</td>
</tr>
<tr>
<td>[Tang et al. 2015]</td>
<td>2015</td>
<td>Mesh</td>
<td>93.16% 93.33%</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>2016</td>
<td>Mesh</td>
<td>97.96% 95.28%</td>
</tr>
</tbody>
</table>

mance in recognition.

5.2.5.4 3D face recognition on Bosphorus database

(1) Experiments with pose and occlusion variations

Comparing to the 3D face scans in FRGCv2 database, the variations challenges of face in Bosphorus database are well classified and labelled. This experiment condition permits us to better study the robustness of our proposed feature to the variations challenges from a easier level to a harder level. In the presentation of our proposed PCM-meshSIFT feature descriptor, the function of the assignment of the canonical direction is to guarantee the algorithm’s stability to overcome the head pose changes. Meanwhile, the keypoint-wise feature description and matching are suitable to identify the samples with external occlusion. Based on these characteristics of PCM-meshSIFT feature descriptor, we perform the experiments on partial Bosphorus database, i.e. the probe sets consists of neutral expressive face, face with various rotations and occlusions. Tab.5.5 lists the rank-one recognition rate in different subsets of Bosphorus DB.

From the viewpoint of the effectiveness of fusion operation, the combination of three feature descriptor enhance the recognition performance and the improve the robustness of algorithm to variations. Observing the results achieved in different cases, the recognition rates with the mean fusion rule are higher than using the individual feature descriptor. Especially, when the face rotates 90° in two directions, the self-occlusion degrades seriously the discriminative power of three feature descriptors. However, the fusion techniques combines the shape information represented by different geometric attributes, enrich the informativeness of feature and improve the recognition performance.

Moreover, according to the face recognition rates involving pose changes, our
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Table 5.5: Rank-one recognition rate with pose and occlusion variations on Bosphorus database.

<table>
<thead>
<tr>
<th></th>
<th>Baseline (194 scans)</th>
<th>HOC1</th>
<th>HOC2</th>
<th>HOC3</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Pose Changes (1365 scans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YR 10°</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>YR 20°</td>
<td>99.05%</td>
<td>98.10%</td>
<td>97.14%</td>
<td>99.05%</td>
<td></td>
</tr>
<tr>
<td>YR 30°</td>
<td>97.14%</td>
<td>98.10%</td>
<td>94.29%</td>
<td>99.05%</td>
<td></td>
</tr>
<tr>
<td>YR 45°</td>
<td>96.19%</td>
<td>94.29%</td>
<td>92.86%</td>
<td>98.57%</td>
<td></td>
</tr>
<tr>
<td>YR 90°</td>
<td>49.67%</td>
<td>36.19%</td>
<td>20.00%</td>
<td>57.14%</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>89.96%</td>
<td>87.62%</td>
<td>83.59%</td>
<td>92.75%</td>
<td></td>
</tr>
<tr>
<td>Occlusions (381 scans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HandE</td>
<td>98.10%</td>
<td>97.14%</td>
<td>94.29%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>HandM</td>
<td>100%</td>
<td>99.05%</td>
<td>97.14%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Glasses</td>
<td>96.15%</td>
<td>94.23%</td>
<td>88.46%</td>
<td>97.12%</td>
<td></td>
</tr>
<tr>
<td>Hair</td>
<td>92.54%</td>
<td>86.57%</td>
<td>76.12%</td>
<td>95.52%</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>97.11%</td>
<td>95.01%</td>
<td>90.29%</td>
<td>98.43%</td>
<td></td>
</tr>
<tr>
<td>All meshes except YR 90° (1730 scans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>97.95%</td>
<td>96.88%</td>
<td>94.62%</td>
<td>98.73%</td>
<td></td>
</tr>
<tr>
<td>All meshes (1940 scans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>92.37%</td>
<td>90.31%</td>
<td>86.55%</td>
<td>94.23%</td>
<td></td>
</tr>
</tbody>
</table>
proposed method is robust and effective if the rotation degree is no more than 45°. In these cases, the recognition rates with mean fusion rule are all above 97%. However, as mentioned above, the recognition performance is weakened sharply if there is a large pose change, i.e. 90° rotation. The face samples in this case are regarded as a self-occlusion variation, and half face is missing. It results that the keypoints with meaningful shape information on the hidden half face are lost. Besides, according to the 3D face scanning techniques, the large pose change may also alter the topological structure of triangle mesh, especially around the eye areas and the nose area. This kind of alteration leads to the intensively change of estimation results of principal curvature measures.

Furthermore, when the external occlusion is caused by hand, our method remains its recognition performance. The corresponding recognition rates are around 97% and raised to 100% when we perform the fusion of three feature descriptors. If the subject wears the glasses, these is an accepted impact on the discriminative power, and the recognition rate decreases to 97.12%. However, if there are hairs on captured face model, the recognition accuracy is degraded to 95.52%. The difference between the hand occlusion cases and the latter two cases is the alteration intensity of topological structure of meshed face scan. Especially for the hair occlusion case, the long hair of women changes the topological structure of a majority of facial surface. Moreover, when we calculate the principal curvature measures and vectors of the vertices around the hair part, the Borel subset may contain this hair which affect intensively the estimation correctness. Another interesting finding is that the recognition rate if the mouth is hidden (HandM) is higher than the one if the eye is hidden (HandE). We surmise that the eye surrounding area can provide richer shape information for 3D face recognition.

(2) Comparison with the state-of-the-arts Tab.5.6 compares the proposed method with the state-of-the-art on the subset of pose variations in Bosphorus. Hajati et al. [Hajati et al. 2012] reported an approach using patch geodesic moments which are extracted from the texture image contolled by the corresponding range image. However, the cropping, pose correction and alignment in their method were performed manually. Li et al. [Li et al. 2015] proposed the multi-order surface differential quantities based method, which is robust to expression, pose and occlusion variations. Comparatively, our proposed method achieves equal or better results in most parts of this subset, except the YR 20° case. It shows the robustness of the method to the limited pose changes. However, it is also worthy to notice that the rank-1 recognition rate is less than 60% if there are large pose change, i.e. YR 90° case. The recognition of half face missing and self-occlusion is still problematic and need further investigation.

Tab.5.7 compares our method with the state-of-the-art with respect to occlusion variations on the corresponding subset in Bosphorus database. The learning-based
Table 5.6: Performance comparison on the subset of the Bosphorus dataset with various pose variations.

<table>
<thead>
<tr>
<th></th>
<th>10°</th>
<th>20°</th>
<th>30°</th>
<th>45°</th>
<th>90°</th>
<th>PR</th>
<th>CR</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dibeklioglu et al. 2009</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>79.41%</td>
</tr>
<tr>
<td>Maes et al. 2010</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>85.6%</td>
<td>24.3%</td>
<td>-</td>
<td>-</td>
<td>84.2%</td>
</tr>
<tr>
<td>Hajati et al. 2012</td>
<td>92.3%</td>
<td>88.6%</td>
<td>80.0%</td>
<td>38.6%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>69.1%</td>
</tr>
<tr>
<td>Li et al. 2015</td>
<td>100%</td>
<td>100%</td>
<td>99.1%</td>
<td>97.6%</td>
<td>47.1%</td>
<td>99.5%</td>
<td>99.1%</td>
<td>91.1%</td>
</tr>
<tr>
<td>PCM-meshSIFT</td>
<td>100%</td>
<td>99.1%</td>
<td>99.1%</td>
<td>98.6%</td>
<td>57.1%</td>
<td>99.5%</td>
<td>99.1%</td>
<td>92.7%</td>
</tr>
</tbody>
</table>

Table 5.7: Performance comparison on the subset of the Bosphorus dataset with various occlusion.

<table>
<thead>
<tr>
<th></th>
<th>HandE</th>
<th>HandM</th>
<th>Glasses</th>
<th>Hair</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alyüz et al. 2008</td>
<td>93.6%</td>
<td>93.6%</td>
<td>97.8%</td>
<td>89.6%</td>
<td>93.6%</td>
</tr>
<tr>
<td>Colombo et al. 2011a</td>
<td>91.1%</td>
<td>74.7%</td>
<td>94.2%</td>
<td>90.4%</td>
<td>87.6%</td>
</tr>
<tr>
<td>Drira et al. 2013</td>
<td>97.1%</td>
<td>78.0%</td>
<td>94.2%</td>
<td>81.0%</td>
<td>87.0%</td>
</tr>
<tr>
<td>Li et al. 2015</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>95.5%</td>
<td>99.2%</td>
</tr>
<tr>
<td>PCM-meshSIFT</td>
<td>100%</td>
<td>100%</td>
<td>97.12%</td>
<td>95.5%</td>
<td>98.4%</td>
</tr>
</tbody>
</table>

approaches adopted in [Colombo et al. 2011a] and [Drira et al. 2013] detect and restore partially occluded face parts using a subspace of 3D face scans trained with an additional dataset. Both training 3D face scan and probe scans need to be correctly normalized. Our proposed method, without any prior training and alignment, outperforms these methods. But, compared to the approach proposed in [Li et al. 2015], our method is a bit weaker in occlusion involved recognition scenario. Their feature-level fusion method can better control the variations occurred by glasses.

5.2.5.5 Times cost

In order to analyse the computation time of our proposed framework, we evaluate the time cost step by step on a PC with Intel Core i7-3770 CPU (8 cores) @ 3.40GHz with 16GB RAM. The testing samples come from FRGCv2 database, each of which contains 30K vertices and 500 keypoints in average. This test estimates 3 stages in the PCM-meshSIFT based 3D FR system: 3D keypoint detection, 3D keypoint description, and 3D keypoint matching. Specifically, the keypoint detection consists of the generation of Gaussian scale space, the estimation of principal curvature measures in each scale space and the keypoint location. The keypoint description includes the assignment of the canonical direction and the feature configuration. The matching is to match one sample to the whole gallery set containing 466 samples. In particular, because we want to test the time cost in online part, the time
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Table 5.8: Computation time consumed by each step in PCM-meshSIFT based 3D face recognition system

<table>
<thead>
<tr>
<th>Processing Step</th>
<th>Keypoint Detection</th>
<th>Keypoint Description</th>
<th>Keypoint Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Cost</td>
<td>Gaus. Space</td>
<td>λ, Est.</td>
<td>Location</td>
</tr>
<tr>
<td>3.6s</td>
<td>2.2s</td>
<td>0.8s</td>
<td>1.1s</td>
</tr>
</tbody>
</table>

consumed by building the dictionary of gallery set does not account in the matching stage. The detail of the times cost is listed in Tab. 5.8. Specifically, Tab. 5.8 gives the estimated time cost to generate the feature descriptor from a single 3D face scan, and match it to the whole gallery set. The total consuming time is 8.72s, and 76% of the time is occupied by the keypoint detection stage. Especially, to build the Gaussian scale space and to estimate the principal curvature measures/vectors are the most time consuming portions. The first reason is that the coordinate update and the geometric quantities estimation are both performed on each vertex in meshed surface. The consumed time is proportional to the number of vertices. The second reason is that the time expended by these two mentioned processing step is multiplied by the number of the level in the scale space. From this viewpoint, it is easier to conclude the significance and the necessity of the keypoints, which will save too much time in feature extraction and matching.

5.3 Heterogeneous Cross-Resolution related Experiments

In this history of 3D face recognition, the related approaches prefer to adopt the face models in high-resolution. Because the facial surface can be regarded as the convex geometry surface, this kind of 3D models consisting of massive vertices and facets can achieve a higher approximation to the real facial surface. This kind of models are generally used in many 3D face recognition system and database, e.g. FRGC [Phillips et al. 2005], Bosphorus [Alyüz et al. 2008], BU-3DFE [Yin et al. 2006]. In the early years, the devices to capture such kind of data, e.g. Minolta VIVID 910, may take a few seconds for a single session, and during this period, the face is required to keep still, which is not practical for the real-life application scenarios. Fortunately, with the development in both hardware and software of 3D scanning technology, 3dMD and Artec3D are able to provide dynamic flows of 3D face scans at the rate of tens of frames per second. However, these devices are quite expensive and in big size, which are not convenient to operate in practical conditions. Meanwhile, with the increase of quantity of vertices and facets captured
and recorded in one 3D face model, the cost of the data acquisition and the feature calculation also have an explosive growth. Recently, the advent of low-cost and real-time 3D scanning devices, especially the Microsoft Kinect and Asus Xtion Pro Live, make it possible to collect and exploit 3D data in the daily life. The smaller size and the consumer-level price provides these 3D scanning devices a competitive predomination in 3D face recognition.

Due to the popularization of Kinect sensors, low-cost 3D data recording RGB-D information have received increasing attention in the academia in various aspect, namely action recognition, object detection, scene classification, etc. But in face recognition, it is more challenging to adopt such low-cost 3D face data, because, in consideration of the balance of the price against the precision, more noise exists in the 3D face data captured by Kinect. Moreover, the limited resolution of the range image capture by Kinect also set the restriction to generate satisfactory quantity of vertices and facets for building the triangular mesh model. The unqualified 3D meshed face model certainly leads to a weak recognition performance of face recognition system.

Based on this condition, this thesis proposes to combine the high-resolution samples and the low-resolution samples to perform a heterogeneous cross-resolution face recognition, which combines the advantages of both kinds of face samples. In this heterogeneous cross-resolution face recognition framework, all the high-resolution samples are registered into the gallery set, while all the low-resolution sample are assigned into the probe set. The apparent merit of this recognition pattern is to separate the most time-consuming part, i.e. the processing of high-resolution samples, and to implement it in the off-line operation. But this high-resolution samples can offer rich shape information during the feature matching step. On the contrary, the data acquisition and the feature extraction steps of the low-resolution are performed in the on-line part, which is more efficient.

In this heterogeneous cross-resolution face recognition, we continue to use the PCM-meshSIFT feature descriptor, which has been proven effectiveness in 3D face recognition and robust to uncontrollable variations in previous section. Furthermore, the adjustable region size of Borel subset for estimating the principal curvature measures is helpful to build the facial features in the facial samples in different resolutions. However, due to the low-quality of the low-resolution 3D face data used in this approach, a Signed Distance Function (SDF) base optimization method is applied to enhance the quality of facial surfaces. Meanwhile, we replace the keypoints detected via Difference of Curvature based method by the landmarks, because the DoC-based keypoint detection method cost too much time. We take the Active Shape Models based landmark location method to define the initial anchor points, and locates more keypoints based on them. This methods spends less time and makes it possible to use the cross-resolution 3D face recognition in the practical
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

In the following part, we present firstly the cross-resolution 3D face database, named as Lock3DFace database, applied in our approach. Then, according to the low-resolution 3D face scans in Lock3DFace database, we introduce the SDF based optimization method, and the Active Shape Models (ASM) based landmark location mechanism. Finally, we present the experiments performed in this research.

5.3.1 Lock3DFace database

For constructing the Lock3DFace database, 509 subjects (the majority of subjects are college students) participate in the data collection, in which 377 are male and 122 are female. The age distribution of the subject is in the range of 16 to 36 years old. As shown in Fig. 5.10, all the major challenges in face recognition are considered, including expression, pose and occlusion variations. For each subject, at least two video clips are recorded including the samples in neutral expression and frontal pose, and the samples with expression, pose and occlusion variations.

The Lock3DFace database consists of two subsets: the low-resolution face subset and the high-resolution face subset. The low-resolution faces in first subset are acquired by using Kinect V2, the second generation of the Kinect sensor. Similar to the original Kinect, the sensor continues to use infra-red to read the environmental data, but presents higher accuracy over its predecessor. Kinect V2 updates the 2D camera to a higher resolution one that can be used for color video recording. Moreover, it has an increased filed of view, thus reducing the amount of distance.

Figure 5.10: Illustration of samples in Lock3DFace database. (a) neutral, (b)~(f) expression variations, (g)~(h) occlusion, (i)~(l) pose changes.
needed between the user and the sensor for optimal configuration. During the data collection, three types of modalities, i.e. color, depth and infra-red data, are collected in three individual channels in parallel. Specifically, the color images are recorded as the size of 1920 × 1080, and the depth and infra-red images are of the resolution of 512 × 424. In the recorded video clip, there are 60 frames per second.

This subset is introduced in [Zhang et al. 2016b].

The high-resolution faces in second subset are acquired with the structure-light-based ZW-121 3D face scanner developed by Wisesoft. The data collection speed of the capture device is 0.3 s per scan. The scanning precision is less than 0.1 mm and the average distance between the sample points is 0.45~0.55 mm. In the scanned bust of subject, there are approximately 65K vertices and 130K facets.

In this subset, a corresponding texture image of subject is captured by camera in parallel. This subset has not been published yet.

5.3.2 Signed Distance Function based facial surface optimization

As we state above, the captured 3D face range image is too noisy to extract the feature and perform the matching. Although one publication [Li et al. 2013] presents a 3D face recognition framework with one single low-cost 3D face frame to achieve good performance, the method requires a big gallery set and the samples of each subject enrolled are expected to carry different variations, which is not always fulfilled in the real-life condition. Fortunately, in the low-resolution face subset of Lock3DFace database, a video clip containing a series of frames is offered for each sample. With combining several frames in one sample, we can reconstruct a 3D face model in better quality via 3D accumulation and refining techniques [Hernandez et al. 2012, Izadi et al. 2011].

In our approach, we reconstruct and optimize the 3D face scan with Signed Distance Function. In [Curless & Levoy 1996], this volumetric non-parametric surface representation method is introduced to combine the depth maps of cropped faces. The core function in this algorithm is called Signed Distance Function (SDF), which will iteratively optimize the position of point on surface according to the new incoming depth images.

Specifically, the algorithm adopts a continuous implicit function $D(p)$. The function is the weighted distance of each point $p$ in the spatial space to the nearest facial surface along the line of sight to the sensor. The signed distance is further defined following this rule: If the point $p$ is between the sensor and the facial surface, the distance is positive. If the point correctly locates on the surface, the distance is 0. Otherwise, the distance is negative. We construct the Signed Distance Function by combining the signed distance $d_1(p), d_2(p), \cdots d_n(p)$ and weight function $w_1(p)$.

\footnote{www.wisesoft.com.cn}
3.3 Landmarks location with modified Active Shape Models

In the PCM-meshSIFT based 3D face recognition framework, the keypoints are located automatically. The location method is to build the Gaussian scale space
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Figure 5.12: Illustration of reconstructed 3D face model with $n$ depth images (frames), $n$ is shown in the last row.

with principal curvature measures, and find the extreme of Difference of Curvature to become the keypoints. The advantage of this method is the keypoint offers the scale invariance to the 3D facial feature, while the drawback of this keypoint detection is the high time cost. Because the coordinates of each vertices are required to update for each scale in Gaussian scale space, and the principal curvature measures are also needed to estimated in each smoothed surface. In order to makes the application-wise cross-resolution 3D face recognition approach more efficient, we propose to locate the landmarks on 2D texture images, and generate automatically more keypoints based on these landmarks.

In our approach, we use Active Shape Models to locate the initial landmarks. Active shape models are statistical models of the shape of objects which iteratively deform to fit to an example of the object in a new image [Cootes et al. 1995]. The shapes are constrained by the PDM (point distribution model) Statistical Shape Model to vary only in ways seen in a training set of labelled examples. The shape of an object is represented by a set of points (controlled by the shape model). The ASM algorithm aims to match the model to a new image. The ASM works by alternating the following steps: (1) Generate a suggested shape by looking in the image around each point for a better position for the point. This is commonly done using what is called a ”profile model”, which searches strong edges or uses the Mahalanobis distance to match a model template for the point. (2) Conform the suggested shape to the point distribution model, commonly called a ”shape model” in this context.
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

As presented in Section 5.3.1, a 2D texture image is collected in both the low-resolution face subset and the high-resolution face subset. In this texture image, the Active Shape Models based landmark location method [Milborrow 2009, Milborrow 2007] is adopted to locate 77 landmarks on these 2D texture images, as shown in Fig. 5.13(a). These landmarks distribute in the typical facial areas, e.g. eye corners, eye outline, eyebrows, nose tip, nose wings, nose lower corner, mouth outline, mouth corners, and the face outline. Although these landmarks are typical texture-based interest points, they do not cover all the typical shape-based interest region on face. According to the distribution of principal curvature measures shown in Fig. 5.2 and the distribution of keypoints shown in Fig. 5.3, the shape information characterized by the keypoints locating on the forehead region and the cheeks regions is equivalently meaningful and discriminative. Therefore, by regarding the initial located landmarks as anchor points (the blue points in Fig. 5.13(b)), we assign 40 derivative keypoints (the red points in Fig. 5.13(b) which are the midpoints, trisection points or quarter points of certain lines (the green lines in Fig. 5.13(b)) between some anchor points. On the contrary, the initial landmarks locating in the boundary of face may be lost after the face cropping. Furthermore, the topological structure of triangular mesh grid around this kind of landmarks is unstable, which easily results in the incorrect estimation of principal curvature measures associated to these landmarks. Thereby, we finally eliminate this kind of landmarks (the orange points in Fig. 5.13(c)). Particularly, this type of landmarks are also mapped to the 3D face model for cropping the region of interest. Besides, some landmarks are too close to generate different 3D facial features, we hence merge them as a single keypoint (the green points in Fig. 5.13(c)).

In summary, the 2D face texture image based keypoint location consists of three steps: (1) Location of initial 77 landmarks (anchor points) by ASM; (2) Generation of 40 new keypoints based on the anchor points; (3) Elimination of 16 landmarks over the face boundary and merging the 5 pairs of close landmarks. Finally, we confirm 96 keypoints on each face scans, including the low-resolution face scan and the high-resolution face scan.

In addition, due to the replacement of keypoint detection method, we can not obtain the most suitable scale $\sigma_s$ for each keypoint, which is generated in the building of Gaussian scale space. Thereby, in the PCM-meshSIFT feature extraction process, we set the neighbourhood areas with a radius as 15mm for assigning the canonical direction, and for building the HOC feature, two radii used in neighbourhood area (i.e. daisy flower pattern) for building the HOC feature are set as 10mm and 6mm respectively.
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Figure 5.13: The keypoints distribute on the 2D face texture image. (a) Landmarks (blue points) located by Active Shape Model method; (b) Derivative keypoints (red points) located depending on the landmarks (blue points); (c) Eliminate the boundary landmarks (orange points) and combine the close points (green points).

5.3.4 Experiments

The experiment performed in this section consists of two parts: 3D face recognition with homogeneous resolution data, and 3D face recognition with heterogeneous resolution data. In the first part, the identification performance is evaluated on the 3D face scans in same resolution (the high-resolution data and the low-resolution data respectively). They are the performance baselines of our proposed method on Lock3DFace database. In the second part, we examine the recognition performance carried on heterogeneous cross-resolution face data. After that, we give the time cost of each processing step in the 3D face recognition framework.

5.3.4.1 Experiment Settings

In the Lock3DFace database, there are 509 subjects with 4 samples in neutral expression, in which 2 samples are in high-resolution and 2 samples are in low-resolution. In the experiment with homogeneous resolution data, the first neutral face is registered as gallery sample, while another one is the probe sample. The experiments with high-resolution data and low-resolution data follow this same setting. In the experiment with heterogeneous resolution data, as aforementioned, the first neutral face in high-resolution is enrolled as gallery sample, and the second neutral face in low-resolution is the probe sample.

All the face scans in both resolution are preprocessed following spike and noise removing, and hole filling. With the Active Shape Models based keypoint location method presented in Section 5.3.3, 96 keypoints are assigned in the 2D texture image and mapped on the 3D meshed face model. Particularly, although 15 landmarks
locating on the outline of face do not count as the final keypoints, they are also mapped on the 3D face model to crop the face region. Due to the lack of scale invariance of feature descriptor, the face scans are normalized to the same size by calculating the transformation matrix with the positions of five landmarks, \textit{i.e.} the nose tip, the eye corners and the mouth corners.

In the estimation of principal curvature measures, we define a group of region sizes of Borel subset. Moreover, the groups of region sizes are different for the face samples in different resolutions. In the heterogeneous resolutions related experiments, there will be a region size related one-to-one recognition scenario. It is expected to evaluate the most suitable region size for samples in different resolutions. Specifically, the examined regions sizes for high-resolution samples are the Borel subsets of five sizes, \textit{i.e.} 3, 4, 5, 6 and 7 -ring, while for low-resolution samples, there are another group of sizes of Borel subsets, \textit{i.e.} 1, 2, 3, 4 and 5 -ring.

5.3.4.2 3D face recognition with homogeneous resolution data

In similar to the experiments performed in Section 5.2.5, the recognition performance is tested with three individual facial feature descriptors, and then the matching scores are fused following score-level mean fusion rule to achieve the fusion related recognition rates. In parallel, five region sizes of Borel subsets are evaluated. In Tab.5.9 and Tab.5.10, the experimental results evaluated with the high-resolution samples and the low-resolutions samples are listed respectively.

From the results achieved in two scenarios, the fusion method also enhances apparently the discriminative power of the individual feature descriptor. In the experimental scenario with high-resolution face samples, the fusion method increases averagely the recognition rates around by 5%. And in the scenario with low-resolution face samples, the accuracy is averagely raised around by 7%. Furthermore, there is some changes between the recognition rates achieved in different region sizes. For the high-resolution based scenario, the feature descriptor associated to 5-ring Borel subset achieves the best performance. For the low-resolution based scenario, the highest recognition rate is obtained by using the Borel subset with 3-ring Borel subset to estimate the principal curvature measures and build the feature descriptor.

Comparing globally the experimental results listed in two tables, the proposed feature performs better in the high-resolution scenario. In consideration of the results obtained with fusion strategy, the accuracy of the high-resolution scenario is about 5% higher than the low-resolution. The disparity of the quality of 3D face scan is the main cause. Although we have optimized the facial surface of the low-resolution samples, the disparity between the scanned facial surface and the real facial surface still enlarges the intra-class distance. Furthermore, the reduction
Table 5.9: The rank-one recognition rates with the high-resolution samples in Lock3DFace database

<table>
<thead>
<tr>
<th></th>
<th>3-ring</th>
<th>4-ring</th>
<th>5-ring</th>
<th>6-ring</th>
<th>7-ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOC₁</td>
<td>88.71%</td>
<td>89.90%</td>
<td>90.69%</td>
<td>90.10%</td>
<td>88.12%</td>
</tr>
<tr>
<td>HOC₂</td>
<td>88.32%</td>
<td>90.30%</td>
<td>90.50%</td>
<td>89.70%</td>
<td>87.92%</td>
</tr>
<tr>
<td>HOC₃</td>
<td>86.73%</td>
<td>89.90%</td>
<td>89.31%</td>
<td>88.12%</td>
<td>84.75%</td>
</tr>
<tr>
<td>HOC₁+HOC₂+HOC₃</td>
<td>93.07%</td>
<td>95.84%</td>
<td>96.24%</td>
<td>94.46%</td>
<td>92.67%</td>
</tr>
</tbody>
</table>

Table 5.10: The rank-one recognition rates with the low-resolution samples in Lock3DFace database

<table>
<thead>
<tr>
<th></th>
<th>1-ring</th>
<th>2-ring</th>
<th>3-ring</th>
<th>4-ring</th>
<th>5-ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOC₁</td>
<td>82.77%</td>
<td>85.15%</td>
<td>85.94%</td>
<td>84.75%</td>
<td>82.38%</td>
</tr>
<tr>
<td>HOC₂</td>
<td>82.38%</td>
<td>84.95%</td>
<td>85.35%</td>
<td>83.96%</td>
<td>82.18%</td>
</tr>
<tr>
<td>HOC₃</td>
<td>80.40%</td>
<td>81.98%</td>
<td>82.97%</td>
<td>83.17%</td>
<td>79.60%</td>
</tr>
<tr>
<td>HOC₁+HOC₂+HOC₃</td>
<td>90.69%</td>
<td>92.67%</td>
<td>93.27%</td>
<td>92.28%</td>
<td>90.30%</td>
</tr>
</tbody>
</table>

of the number of vertices and facets also results to the characterization ability of feature descriptor. Because in the same neighbourhood area for generating the HOC feature, there are fewer vertices around the keypoint to build the histogram.

5.3.4.3 3D face recognition with heterogeneous resolutions data

In the heterogeneous resolution based 3D face recognition experiment, all the enrolled samples are the high-resolution face scans, while all the text samples are the low-resolution face scans. For the gallery set and the probe set, we assign different groups of regions scales respectively. The experimental results are listed in Tab.5.11. All the rank-one recognition rates are computed with the score-level fusion following mean rule.

From the listed results, the average of recognition rates between the face scans in different resolution is 77.12%. This recognition accuracy demonstrates that our proposed framework, including the modified PCM-meshed feature description and SRC based matching method, has acceptable performance to match the cross-resolution face scans. But comparing to the homogeneous resolution based experimental scenarios, the identification performance of the feature descriptor reduces sharply. The recognition accuracy decreases averagely by 17.33% (compared to the high-resolution based results) and 14.72% (compared to the low-resolution based results), respectively. Even though we have normalized the histograms of principal curvature measures used in feature extraction, the disparity of the density of the vertices between the scans in two kinds of resolution also has intensive impact on
Chapter 5. 3D Face Recognition in the presence of Pose, Occlusion, and Heterogeneous Cross-resolution Variations

Table 5.11: The rank-one recognition rates of heterogeneous cross-resolution based 3D FR on Lock3DFace database

<table>
<thead>
<tr>
<th>$HOC_1 + HOC_2 + HOC_3$</th>
<th>Low Res.</th>
<th>High Res.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-ring</td>
<td>4-ring</td>
</tr>
<tr>
<td>1-ring</td>
<td>73.07%</td>
<td>76.63%</td>
</tr>
<tr>
<td>2-ring</td>
<td>76.63%</td>
<td>77.62%</td>
</tr>
<tr>
<td>3-ring</td>
<td>78.61%</td>
<td>79.80%</td>
</tr>
<tr>
<td>4-ring</td>
<td>78.61%</td>
<td>79.80%</td>
</tr>
<tr>
<td>5-ring</td>
<td>78.61%</td>
<td>79.80%</td>
</tr>
</tbody>
</table>

the keypoint matching.

From the viewpoint of the region scales of Borel subset, the similar conclusion to the analysis above can be achieved that the high-resolution scans related feature computed in radius 5 and the low-resolution scans related feature computed in radius 3 are able to provide better discriminative power. In this case, we achieve the best recognition rate of 83.37%. In fact, it is not a coincidence to obtain this result, because the spatial areas of the Borel subset are similar by using this combination of the region scales for different resolutions based face scans. It leads to the closer estimation results of each principal curvature measures and the alike distribution of principal curvature measures around each keypoint.

5.3.4.4 Time Cost

To analyse the computation efficiency of our proposed the heterogeneous cross-resolution 3D face recognition system, the consuming time of one 3D face scan in low-resolution are listed in Tab.5.12. The statistics is achieved in the online part, which is the processing performed on the low-resolution face scan. Furthermore, the test of the time cost is estimated on a PC with Intel Core i7-3770 CPU @ 3.40GHz with 16GB RAM. The statistics is reported in three stages: preprocessing, feature extraction, and matching. Specifically, the preprocessing is the SDF-based 3D facial surface reconstruction step. The feature extraction stage consists of 4 portions: the estimation of principal curvature measures, the keypoint location, the canonical direction assignment and the feature description. The keypoint location includes the three steps shown in Fig.5.13. The matching step is to match one test sample to the whole gallery set (509 gallery samples). The total time consuming for one testing face scan is 0.46s. For a practical 3D face recognition application, the efficiency of the framework is promising. The modified keypoint detection method efficiently decreases the time cost. Besides, the most time consuming part is the optimization step, which accounts for 31.7% of total processing time. In this step, 60 frames of range images are merged into the 3D facial surface reconstruction. A possible solution to reduce the time cost in this step is to lessen the number of
frames merged. Because the optimization quality of facial surface increase slower after merged around 40 to 50 frames. In the future work, it’s worthy to test the impact of the number of merging frames on cross-resolution face recognition.

5.4 Conclusion

In this chapter, we proposed an efficient 3D face recognition approach based on principal curvature measures and principal curvature vectors for 3D meshed face scans. To evaluate the effectiveness of generalized curvatures for 3D face recognition, we designed a generalized curvature-based 3D face recognition pipeline based on the general 3D keypoint detection, description and matching framework. In particular, 3D keypoints are detected by using all the three principal curvature measures. 3D keypoint descriptors are constructed by using all the three principal curvature measures as well as the principal curvature vectors. Finally, SRC-based fine-grained matching strategy is employed for 3D face matching. To combine the description powers of all three curvatures, score-level fusion scheme is used to achieve the final decision score. In our experimental parts, we first evaluate our method on the whole FRGC v2.0 database and archived an identification rate of 98.17%. Experimental results demonstrated that the these three principal curvature measures offer significant complementary information for 3D face description and recognition, and the score-level fusion of multi-scales and multi-features enhances the recognition performance following mean rule. In addition, the PCM-meshSIFT based 3D face recognition system achieves the considerable recognition accuracy in the experiments performed on pose invariants and occlusion subsets in Bosphorus database. The corresponding results demonstrate that our proposed method is effective to handle the pose change under 45° rotation and the partial occlusions. However, if there exists extensive missing face part, the discriminative power of the method decreases sharply.

In order to deal with the heterogeneous cross-resolution based 3D face recognition utilized in some practical application, we propose to register the high-resolution face samples in gallery set and test the low-resolution face samples in online testing stage. The Signed Distance Function is adopted to reconstruct the low-quality fa-
cial surface by merging 60 frames of range images captured by Kinect V2. And the
keypoint location method is replaced by a Active Shape Models based 2D texture
face image landmarks detection method. More keypoints are determined based on
these detected landmarks. We also explore the recognition performance of PCM-
meshSIFT feature descriptor on homogeneous resolution data and heterogeneous
resolution data in Lock3DFace database. The experiment results show that the
feature descriptor is effective to identify the subject in both high-resolution and
low-resolution. Even though the disparity of the heterogeneous sample in differ-
ent resolutions enlarges the intra-class distance of subject, this modified PCM-
meshSIFT based 3D face recognition can efficiently achieve the successful matching
rate of 83.37% in this experimental scenario.
Shape Analysis based Anti-spoofing 3D Face Recognition with Mask Attacks

6.1 Introduction

Due to the natural, non-intrusive and contactless acquisition pattern of facial data, face recognition (FR) technology has been accepted extensively by public and applied widely in our daily life [Jain et al. 2004]. This effective and reliable biometrics recognition pattern has spread in numerous practical applications like time attendance, bank service, access control, boarder entry & exit, criminal investigation, etc. The general goal of face recognition is to identify correctly and stably the person via their face, including the 2D face texture image and 3D face model. When applying the face recognition technology in the real-life application, there are more environmental variations increasing the difficulty of face recognition, such as the head pose change, the variation of facial expression, the external occlusion, the missing of partial face, etc. Therefore, some more supplementary goals are proposed for face recognition technology to become more robust to these unexpected variations. In recent years, massive researches and achievements towards face recognition via 2D facial images and 3D facial scans have been...
Chapter 6. Shape Analysis based Anti-spoofing 3D Face Recognition with Mask Attacks

studied and reported for enhancing and improving the recognition performance of face recognition system [Bowyer et al. 2006, Mian et al. 2007, Mian et al. 2008, Bronstein et al. 2003, Li et al. 2015, Huang et al. 2012b, Tang et al. 2015]. Meanwhile, in the previous two chapters, this thesis also propose two 3D face recognition methods, i.e. LPCMP-based 3D face recognition framework and PCM-meshSIFT based 3D face recognition approach, for achieving stronger discriminative power and becoming more robust to the undesirable variations.

However, a potential menace hiding behind these successful researches, called spoofing attacks, threats insidiously the security of face recognition system. Spoofing attack is defined as an intrusive act of deceiving a biometric system by presenting a fake evidence or copied biometric trait to obtain valid authentication [Nixon et al. 2008]. Particularly, the fake evidence commonly presented in face spoofing attacks is a photograph or a video of the valid user, which can be captured at distance or collected via internet. Due to the development of 3D scanning and printing techniques, the convenient manufacture of 3D mask provides an easier way to intruders for masquerading as a registered person in FR system. Moreover, a social public website “Thats My Face” \(^1\) provides the manufacturing service of wearable 3D mask, which only demands to offer one frontal photo (a side-view photo demanded as option). Unfortunately, the convenient facial data acquisition and the facial sample manufacture, which should be the advantage of face recognition, become gradually the jeopardy and the calamity to face recognition system.

Between the fraud mask and genuine face, the shape dissimilarity can also be the clue to distinguishing them. Especially, the difference of the geometric shape in local region between the mask and the genuine face can be highlighted by the local geometry-shape based facial feature descriptor. In consideration of the outstanding performance of PCM-meshSIFT facial feature descriptor performing on the genuine face samples, in this chapter, we evaluate its anti-spoofing performance on Morpho database. The framework and the detail of the this facial feature descriptor is presented extensively in Chapter 5. Furthermore, during the practical face recognition application, the probability of spoofing attacks is quite lower than the general verification cases of genuine faces. Therefore, we also design a series of anti-spoofing evaluation experiments for simulating a real-world case. We extend the scale of database by adding FRGCv2 database to increase massively the quantity of genuine faces, and test the performance of the 3D facial feature descriptor in this real-world simulating case.

\(^1\)http://www.thatsmyface.com
6.2 Related Work

To deal with the vulnerability of face recognition system mentioned above, many effective countermeasures have been proposed to withstand the spoofing attacks. Liveness detection [Pan et al. 2007, Chetty & Wagner 2006, Kollreider et al. 2008, Zhang et al. 2011], motion detection [Kollreider et al. 2005, De Marsico et al. 2012] and texture analysis [Li et al. 2004, Määttä et al. 2011] are three principal categories of anti-spoofing approaches [Erdogmus & Marcel 2013, Erdogmus & Marcel 2014] against the spoofing attacks by using photos and videos. Moreover, Morpho database\(^2\) and 3DMAD database [Erdogmus & Marcel 2013], which contains 2D, 2.5D and 3D face samples of genuine facial subjects and fraud 3D mask worn by “impersoners”, are introduced to evaluate the anti-spoofing performance of the countermeasures against mask attacks. In the following part, some representative anti-spoofing approaches are introduced as follows.

Due to the difference of reflectance between real face skin and materials used to manufacture masks, the work of Kim et al. [Kim et al. 2009] aimed to analyse the distribution of albedo values for illumination at various wavelengths. Based on Fisher's linear discriminant, they selected a 2D feature vector consisting of radiance measurement to be the classification criteria in visual and NIR spectrum (685 and 850 nm respectively). Similarly, Zhang et al. published their mask detection approach based on multi-spectral analysis in [Zhang et al. 2011]. They claimed to abandon visual face image, but to analyse multi-spectral images captured in two discriminative illuminations whose wavelengths are 850 nm and 1450 nm. They measured the albedo curves of different materials and trained SVM classifier for distinguishing genuine faces and fraud mask. Even though these two papers are effective in mask detection with optical methods, an extra multi-spectral capture device is obligatory in their countermeasures which increases the time-cost and the complexity of face recognition system.

Lately, Kose et al. reported their anti-spoofing works based on texture and depth information extracted in 2D and 2.5D images from Morpho Database. Three baseline face recognition algorithms are tested and an experimental scenario associated to mask attacks is mentioned in [Kose & Dugelay 2013b]. Furthermore, they describe the texture and depth facial images with LBP features and train the linear SVM classifier to determine whether the input sample is genuine or fake in [Kose & Dugelay 2013a]. In order to synthesise the advantages of the features in texture and depth images against mask attack, a score-level fusion of them were proposed and examined in [Kose & Dugelay 2013c]. They concludes that texture analysis is an effective method for developing such a countermeasure. Similarly, Erdogmus et al. introduced their 3D Mask At-

\(^2\)http://www.morpho.com
tack Database (3DMAD) and anti-spoofing countermeasures based on three extended LBP algorithms in [Erdoğmus & Marcel 2013]. Some more comparative anti-spoofing experimental results on Morpho and 3DMAD databases were reported in [Erdoğmus & Marcel 2014]. These approaches mentioned above have proved that the photometric features in 2D facial images are capable to protect the system from mask attacks.

However, even though two features based on 3D face scans are used to perform the anti-spoofing test, including Thin Plate Spline (TPS) warping parameters and Iterative Closest Point (ICP), the anti-spoofing performance of the local features based on 3D shape analysis has not been deeply studied yet. Moreover, to deal with such a practical problem associated to real-world case, all the evaluation results and conclusions are obtained based on the experiment with extreme conditions, which means that the majority of probe set is fraud mask samples. For optimizing the anti-spoofing countermeasures and the evaluation scenarios, in this chapter, we evaluate the anti-spoofing performance of PCM-meshSIFT feature descriptor in both the experimental case with extreme testing condition and the real-world simulating case.

6.3 Experiments

6.3.1 Database

In the series of experiments for examination of verification ability and evaluation of anti-spoofing performance, Morpho database and FRGC v2.0 database are applied in this section. The FRGCv2 database has been presented, please refers to Section 4.3.1. The Morpho database contains both the genuine face samples and the fraud mask samples. In Morpho database, 16 masks were manufactured according to 16 persons. Their shape information is captured by 3D scanner with structured light, and the mask is manufactured with 3D printer by Sculpteo 3D Printing [Kose & Dugelay 2013a]. The samples in Morpho database are divided into two categories: (a) 10 genuine face samples are captured for each one of 20 subjects; (b) each one of 20 subjects wearing his/her or other’s mask are captured for around 10 times. Specifically, there are two cases for the samples belonging to (b). The person wearing his/her own mask is marked as type $A_A$ sample, otherwise it’s marked as type $A_B$ sample. In the following experiments, type $A_A$ samples and type $A_B$ samples both count as spoofing mask attacks. Fig.6.1 shows some examples of genuine face samples, including type $A_A$ and type $A_B$ spoofing samples.
6.3.2 Experiment Scenarios

The basic target of one 3D face recognition system with common genuine face is to verify the identity correctly, and the supplementary target is to distinguish fraud masks from genuine faces. Therefore, we firstly define a series of thresholds of similarity measurements $t^i_t$ for achieving a corresponding set of verification rate in a baseline scenario, named Sc-Base. Then we apply each corresponding threshold of similarity measurement to evaluate the anti-spoofing performance with only fraud masks as probe set in scenarios named Sc-Attk. Finally, we perform the real-world simulating scenario, named Sc-RW, by increasing massively the number of genuine face in probe set. The details of experimental configurations of scenarios are listed in Table 6.1.

In Table 6.1, several symbols need to be specified more clearly. $G_M$ and $S_M$ refers respectively to all genuine face scans and all spoofing mask scans (including type $A_A$ and type $A_B$ mask scans) in Morpho database, similar to $G_F$ refering to FRGC database. $G_M^1$ and $G_M^i$ represents respectively the set of the first scan and others of each subject in $G_M$, which is similar for $G_F$. $S_M^p$ refers to a partial set of $S_M$, which makes the ratio of fraud masks to genuine faces to be 1:99 in Sc-RW-3, while the ratio is 5:95 in Sc-RW-2.
6.3.3 Experimental Results and Analysis

In Sc-Base, we evaluate the baseline of verification performance with True Accept Rate (TAR). Because TAR varies along with different $t_i^r$, in Table 6.2, TAR is presented under several cases with typical False Accept Rate (FAR) decided by different $t_i^r$. TAR is shown in Table 6.2 under several cases in different False Accept Rate (FAR). For scenarios Sc-Attk in Table 6.2, only the spoofing samples perform the probe set for evaluating the anti-spoofing performance. Moreover, TAR and FRR, which are commonly used as criterion for normal verification experiment, can’t be estimated in this case. Because all the testing samples in Sc-Anti are illegal samples with mask and should be rejected, there doesn’t exist the truly accepted samples with mask and the falsely rejected samples with mask. Therefore we evaluate the performance with the criterion named Spoofing True Reject Rate (STRR), which is an special TRR for spoofing scans in Sc-Attk. Spoofing True Reject indicates that the samples with mask is correctly reject by the system. Furthermore, according to the configuration of each scenario, the $t_i^r$ used in Sc-Attk-1 corresponds the one implemented in Sc-Base-1, while the $t_i^r$ of Sc-Attk-1 and Sc-Attk-2 follow the sets in Sc-Base-2.

As shown in Table 6.2, the verification rate of Sc-Base-1 is above 92% except the case when FAR is 0.01. And in Sc-Base-2, the verification rate is 91.98% even FAR equals only 0.001 when the scale of database is extended by adding FRGC
If we compare the discriminative power between these two scenarios, the TAR increases from 84.74% to 96.38% if FAR is 0.01, which means that if more samples enrolled in the system, the performance can be evaluated more properly. However, in Sc-Base-1, the case FAR=0.01 is particular which needs to be clarified. Because the scale of genuine face scans in this scenario is 177, which means there are only 1 or 2 samples accepted falsely in this case, the verification rate that TAR is 84.75% is also acceptable. It is easy to understand the blanks in Sc-Base-1 and Sc-Attk-1 when FAR is 0.001. Based on the contrary thought, the results in first row remain blank in the scenarios involving FRGC database. Because there are 3,721 genuine face scans in such scenarios, and FAR equals 0.1 means that there are 370 samples accepted falsely which can’t show the performance correctly. Therefore, according to the results of verification baseline examination, we can conclude that our 3D facial geometric attribute based feature description guarantees the verification power for genuine faces.

In scenario Sc-Attk-1, STRR is above 62.82% and arrives 75.64% when FAR decreases to 0.01. Apparently, STRR arises with the decrease of FAR due to the more restricted threshold $t^i_\mu$. After extended the database in Sc-Attk-2 and 3, STRR is higher than 62.97%, then reaches 81.28% and 84.59% respectively with different scale of spoofing probe set. In this case, STRR is lower a bit with same FAR than Sc-Attk-1, because the scale of gallery set is larger and $t^i_\mu$ is more rough. This evaluation results demonstrate that our proposed PCM-meshSIFT-based FR framework can complete the anti-spoofing mission against mask attacks.

Furthermore, for simulating the real-world testing case comprehensively, we implement the probe set uniting genuine faces from both databases and the spoofing probe set (i.e. $S_M$). Fig.6.2 shows the Detection Error Tradeoff graph (DET) to present the experimental results. Apparently, in DET graph, the lower Equal Error Rate (EER) means the better evaluated verification result of scenario with same gallery set. The upper plot in Fig.6.2 shows the scenarios that the gallery set is limited as Morpho database, as the control group for experiment, and the lower one shows the simulating scenarios with extended database. Remark that the lower one in Fig.6.2 only shows the part when FAR and FRR is lower than 35% for showing more clearly. Moreover, due to the introduction of fraud mask samples in Sc-RW scenarios, it is not proper to evaluate the performance by FAR and FRR. Instead of them, we replace SFAR and SFRR to evaluate it. Specifically, for one subject in the gallery set, only accepting its own genuine face is regarded as correct operation, whilst accepting the genuine faces belonging to other individuals or all the fraud masks are treated as incorrect decision. In the first plot, EER of Sc-RW-1 is 9.3% which is observably higher than the baseline scenario Sc-Base-1. It’s obvious to study that the involvement of spoofing mask samples decreases the verification performance of system. Similarly, EER of Sc-RW-2 and Sc-RW-3 are higher than
Chapter 6. Shape Analysis based Anti-spoofing 3D Face Recognition with Mask Attacks

the corresponding baseline result in Sc-Base-2, but they decline comparatively to Sc-RW-1 as 3.1% and 2.8% respectively. A similar conclusion obtained as before that two goals have been achieved: (1) the verification system guarantees a high verification ability and (2) it possesses the distinguishable power against spoofing attacks in real-world simulating case.

6.3.4 Comparison with the state-of-the-art approaches

In this subsection the comparison with the state-of-the-art approaches using Morpho database is also given in Table.6.3. According to the experimental configuration assigned in [Kose & Dugelay 2013b, Erdogmus & Marcel 2014], here we use scenario Sc-Base-1 to compute the EER. Then adopt the same threshold in Sc-Attk-1 to compute the SFAR, that is, Spoofing False Accept Rate. Spoofing False Accept indicates to the case that the samples with mask is false accepted by the system. In Table 6.3, we only report the corresponding experimental results using 3D meshed facial samples. Comparing to 2D texture image and 2.5D depth image, the face recognition based on 3D face samples generally achieve higher verification performance with all genuine face samples in the test. The warping parameters related method in [Kose & Dugelay 2013b] achieves the lowest EER of 3.85%, which is better than 6.91% EER obtained by our method. However, the WP-related FR system is the most vulnerable one among the reported systems. Our PCM-meshSIFT-based method is the most robust system when replace all the probe samples by the samples with masks. We achieve the lowest SFAR of 33.10% in such experiment. It proves that the minor shape difference between genuine face and manufactured mask can be detected and highlighted by our principal curvature measure based 3D facial feature, and it is effective to enhance the security level of FR system.
Chapter 6. Shape Analysis based Anti-spoofing 3D Face Recognition with Mask Attacks

Table 6.3: Comparison of verification performance with spoofing attacks in Morpho database.

<table>
<thead>
<tr>
<th></th>
<th>Texture Image</th>
<th>Depth Image</th>
<th>3D Mesh Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
</tr>
<tr>
<td>EER</td>
<td>5.90%</td>
<td>6.54%</td>
<td>7.27%</td>
</tr>
<tr>
<td>SFAR</td>
<td>72.87%</td>
<td>59.94%</td>
<td>88.94%</td>
</tr>
</tbody>
</table>

Note: (1) Results reported in [Kose & Dugelay 2013b]; (2) Results reported in [Erdogmus & Marcel 2014]; (3) Results reported with our proposed method.

6.4 Conclusion

In this chapter, we adopt the 3D shape analysis based PCM-meshSIFT method to describe and quantize the facial shape, and further highlight the shape difference between the fraud mask and the genuine faces. Both Morpho database containing fraud mask samples and FRGCv2 database containing massive genuine face samples are utilised to perform the experiments. Specifically, three kinds of experimental scenarios are designed to evaluate the baseline verification performance, to examine the anti-spoofing performance in extreme condition in the real-world simulating case. From the experimental results in these experimental scenario, the shape analysis based local feature descriptor possesses the promising verification ability for genuine face and the distinguishable capability for manufactured masks.
Chapter 7

Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Contents

7.1 Introduction .................................................. 130
  7.1.1 Overview of Hand-dorsa Vein Recognition .......... 130

7.2 Hand-dorsa vein image acquisition .......................... 132

7.3 Framework of multi-source keypoint based hand-dorsa vein recognition ........................................ 134

7.4 Multi-level Keypoint Detection ............................... 136
  7.4.1 The Difference of Gaussian Detector ................. 136
  7.4.2 Harris and Hessian Keypoint Detection ............... 138
  7.4.3 Design of Multi-level Keypoint Detection .......... 139

7.5 Oriented Gradient Maps based Representation ............... 141
  7.5.1 Representation of Complex Neuron Response .......... 141
  7.5.2 Oriented Gradient Maps by Response Vectors ........ 143
  7.5.3 The Properties of Distinctiveness and Invariance .. 143
  7.5.4 Design of the OGM based Keypoint Detector ....... 144

7.6 Local Feature Matching .................................... 145
  7.6.1 SIFT-feature based Matching .......................... 145
  7.6.2 Score Level Fusion .................................... 146
  7.6.3 Illustration of Matching Samples .................... 147

7.7 Experiments .................................................. 147
  7.7.1 Effectiveness of Multi-Level Keypoint Detection .... 149
  7.7.2 Discriminative Power of OGMs ........................ 150
  7.7.3 Impact of Gallery Size ................................ 151
  7.7.4 Verification Validation ............................... 152
  7.7.5 Comparison with the State of the Art ............... 155
  7.7.6 Complementarity of Left and Right Hands .......... 156
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

7.7.7 Evaluation on Generalization Ability . . . . . . . . 156
7.7.8 Complexity Analysis . . . . . . . . . . . . . . . . . . . . . . 159
7.8 Summary and Discussion . . . . . . . . . . . . . . . . . . . . 159

7.1 Introduction

In the previous chapter, the anti-spoofing related approaches aim to find out the difference between the genuine and the fake biometrics evidences. Then certain promising countermeasures are performed to prevent the intrusive action by rejecting the authorisation demand with the fake evidence. But the security hazards still remain in the face recognition system even using the actual countermeasures. Considering some practical applications requiring the high security level, to adopt an accessory biometrics recognition model with natural anti-spoofing ability is a good option to increase the security of the system. Motivated by this idea, in this chapter, we introduce the hand-dorsa vein recognition, which enables the biometrics system to avoid any possible deceitful and intrusive actions. Because the special acquisition pattern of hand-dorsa vein image can only work on the human body alive. When the near infra-red ray lights the back of hand, only the hot blood flowing in the vein vessel is able to absorb most of the near infra-red light, while the epidermis and the dermis of skin reflect the light. Therefore, the hand-dorsa vein vessels are much darker than the surrounding parts in hand-dorsa vein images. Obviously, this liveness detection based data acquisition pattern prevent effectively the spoofing activities to deceive the biometrics recognition system.

However, the hand-dorsa vein image is of low contrast and shows a very sparse subcutaneous vascular network. Comparing to the vein vessel existing in fingers or in hand palm, the construction of hand-dorsa vein is much simpler without abundant minutiae. Therefore, it does not offer sufficient distinctiveness in recognition. In this chapter, we propose a novel approach to hand-dorsa vein recognition through matching local features of multiple sources. In contrast to the current studies only concentrating on the vein vessel network, we also make use of person dependent optical characteristics of the skin and subcutaneous tissue. Evaluated on NCUT database part A and part B, our proposed method shows its effectiveness and generalization ability on person identification and verification.

7.1.1 Overview of Hand-dorsa Vein Recognition

Driven mainly by increasing requirements in public security against terrorist activities, sophisticated crimes, and electronic frauds, biometric solutions have witnessed an accelerated pace of growth in the global market of security over the past several
decades. Recently, the vein has emerged as a new biometric trait for the purpose of people identification, and has received growing attention within the community.

Anatomically, veins are blood carrying vessels interwoven with muscles and bones, and the key function of the vascular system is to supply oxygen to each part of the body. The spatial arrangement of vascular network in the human body is stable and unique, and vein patterns of individuals are different, even between identical twins [Kumar et al. 2009]. In this work, we focus on the vein pattern of the back of the hand (i.e. dorsal hand) because it is distinctly visible, easy to acquire, and efficient to process. As compared with other popular biometric traits, the hand vein has several distinguished merits, in particular the following ones:

- Direct liveness detection. Hand veins are sensed using far or near infra-red lighting to capture the temperature difference between hot blood flow inside vein vessels and the surrounding skin, therefore, they can only be imaged on the live body and the images taken on non-live bodies do not contain their spatial vein arrangement;

- Safety. Blood vessel patterns are hardwired underneath the skin at birth; they are hence much harder for intruders to forge.

The pattern of vein as a biometric trait is relatively recent. It was not presented until 1990 when MacGregor and Welford came up with the system named “vein check” for identification [MacGregor & Welford 1991]. Despite the vast vascular network in the human body, hand veins are favoured for their simplicity in terms of acquisition and processing. In last decades, there exist increasing amount of research works focusing on hand vein recognition using the vein pattern in the palm part [Lin & Fan 2004, Malik & Spaanenburg 2007, Ladoux et al. 2009], the back of the hand [Zhao et al. 2008, Kumar & Prathyusha 2009, Wang et al. 2010a] or fingers [Miura et al. 2004].

Although there have already several attempts on hand vein recognition by adopting holistic techniques, e.g. Principal Component Analysis (PCA) [Heenaye-Mamode Khan et al. 2009], Linear Discriminant Analysis (LDA) [Liu et al. 2012], etc., the changes of viewpoint, lighting intensity, distortion, and occlusion largely impede their development. In contrast, local feature based approaches become dominant due to its robustness to the aforementioned disturbing factors. Most of the methods in the literature follow the framework that first segments the region of interest and the hand subcutaneous vascular network from the hand vein image, and then extracts local geometric features for matching such as the positions and angles of short straight vectors [Cross & Smith 1995], vein minutiae and knuckle shapes [Kumar et al. 2009], endpoints and crossing points [Wang et al. 2006a], dominant points [Lin & Fan 2004], etc. All these methods demonstrate reasonable recognition rates on small databases ranging from 32 [Lin & Fan 2004] to 100 subjects [Kumar et al. 2009] [Kumar & Prathyusha 2009].
However, when regarding the problem of hand-dorsa vein recognition, the above techniques suffer from very limited local features because compared with the palm and the finger part, the number of vein minutiae on the dorsal hand is really few, directly leading to the deficiency in capturing the difference of hand vein networks between subjects. Hand dorsa vein images are mostly sensed by the NIR imaging system, irradiating the hand dorsa with the NIR light. In delivering the vein pattern of the hand dorsa, these images hence also convey the optical properties, i.e. the absorption and scattering speciality, of the skin and subcutaneous tissue which mainly consists of three different layers, namely epidermis, dermis, and hypodermis. The randomly inhomogeneous distribution of blood and various chromophores and pigments produces variations of optical properties of these skin layers that are subject dependent [Bashkatov et al. 2011]. These optical properties are investigated as such in medicine for various purposes, e.g., diagnostics, surgery, therapy. In this thesis, we propose to make full use of these optical properties of the hand dorsa for people identification.

7.2 Hand-dorsa vein image acquisition

Figure 7.1 illustrates the system setup where an LED array lamp is exploited to shine infrared light onto the back of the hand. The incident infrared light can penetrate into the biological tissue with an approximate depth of 3 mm, and the randomly inhomogeneous distribution of blood and various chromophores and pigments produces optical characteristics, i.e. absorption and scattering properties that are subject dependent [Bashkatov et al. 2011]. Since the flow of hot blood inside the vein network generally absorbs and scatters more infrared radiation than the surrounding skin and subcutaneous tissue, its curvilinear structures are imaged through a CCD camera associated with an IR filter where the veins appear darker valleys whereas the surrounding skin and subcutaneous tissue displays a landscape or surface, containing various features, e.g., cliffs, ridges, plateaux, basins etc. (see Fig.7.2). The spectral responses or the variations of these optical attributes of the hand-dorsa skin and the subcutaneous tissue, including in particular the vascular network, are thus perfectly modeled by using differential geometric quantities.

Using such a hardware setup depicted in Fig.7.1, a database of 2,040 dorsal hand vein images of both hands of 102 subjects was built by North China University of Technology in 2010, and it was marked as the NCUT Part A database. In order to make the device more practical, another sensor was proposed by NCUT, in which a trade-off was considered between the expenditure of hardware and the quality of hand vein images. The CCD camera and IR optical filter were substituted, leading to around half reduction in total cost. Another dataset, namely NCUT Part B, was collected through this novel device in 2011. It consists of 2,020 dorsal hand vein

132
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.1: Illustration of the NIR imaging system.

Figure 7.2: Hand-dorsa vein images: (a) from the NCUT Part A dataset and (b) from the NCUT Part B dataset.

images of 101 subjects, each of which owns 20 images; half for the left hand and half for the right hand. In contrast to the NCUT Part A database, NCUT Part B dataset is composed of dorsal hand vein images under different acquisition conditions, and the images are more noisy. Since the vein patterns are best described when the skin on the dorsal hand is taut, a handle was mounted at the bottom of the device to position the hand, and the images were thereby roughly aligned. Fig.7.2 shows samples of NCUT Part A and Part B captured with a resolution of 640 by 480 pixels. There were no major illumination variations, but moderate changes in viewpoint (i.e. differing by rotations as well as translations) still can occur since these images were collected in different periods and environmental situations.

As we can see from Figure 7.2, the pattern of the dorsal hand vein is captured and it appears darker within the NIR image. The widths of these vein profiles
change in the range of 30 to 50 pixels. Even though the vein spatial arrangement is visible, it is not very distinguishable from the surrounding bio-tissue. Furthermore, the number of local features, e.g., endpoints and crossing points, is quite limited and usually varies from 5 to 10, thereby making local feature-based approach questionable for the discriminative power as directly applied to these dorsal hand images. On the other hand, the spectral response of the surrounding skin as well as subcutaneous tissue translates the subject-dependent inhomogeneous compositions of blood and various chromophores and pigments, and their variations are also imaged by the NIR sensor. In interpreting the hand vein image as a landscape or a surface in the same way as retinal images [Morgan 2011], the key features of these variations in absorption and scattering characteristics can be perfectly captured through differential geometric properties, e.g., cliffs, ridges, plateaux, basins, including in particular the valley which corresponds to the vascular network. In order to localize these geometric features and hence increase the number of local features for more distinctiveness, we propose to make use of quantities closely related to differential geometry, namely Harris cornerness and Hessian blobness measurements grouped under the multi-level keypoint detection on the dorsal hand image, and DoG based curvature extrema on a human vision inspired representation, i.e. OGMs. They describe complementary geometric attributes, and we introduce them in the subsequent two sections, respectively.

7.3 Framework of multi-source keypoint based hand-dorsa vein recognition

Specifically, in this section, we propose a novel and effective approach to hand-dorsa vein recognition based on local feature matching based on multi-source keypoint. Unlike the overwhelming majority of state of the art techniques which only focus on the venous network, the proposed method makes full use of discriminative clues as offered by the optical properties of NIR dorsal hand images that cover not only the vein areas but also their surrounding skin and subcutaneous regions. In the same way as the retinal image [Morgan 2011], the optical properties conveyed by NIR dorsal hand images are interpreted as landscapes or surfaces, consisting of geometric features like ridges, valleys, summits, etc. Their properties are comprehensively analyzed using differential geometry quantities, resulting in a set of keypoints of multiple-order gradient cues (from the 1st to 3rd order). More precisely, we introduce the Harris-Laplace detector to characterize the elasticity, i.e., length and angle variations, of the underlying surface, through the cornerness measurement of the 1st order gradients [Mikolajczyk & Schmid 2004]. We then describe the hand-dorsa vein areas which coincide with the valley regions of the underlying
landscape because of their absorption and scattering properties. They are identified using the Hessian-Laplace detector [Mikolajczyk & Schmid 2002] which relies on the blobness measurement of the Hessian matrix of the 2nd order gradients. In order to further thoroughly highlight shape changes, *i.e.* the changes in optical properties, of the whole hand dorsa skin and subcutaneous tissue, we also compute a human vision inspired representation, namely Oriented Gradient Maps (OGMs) [Huang et al. 2012c], of the original image and then identify feature points through the Difference of Gaussian (DoG) [Lowe 2004]. Because OGMs are 1st order gradient based and DoG (an approximation of Laplace of Gaussian, LoG) is 2nd order gradient based, these features are essentially 3rd order gradient based and correspond to the points whose curvatures change most on the surfaces. Finally, the keypoints as detected by the previous process between the hand-dorsa images of the same subject are robustly associated using local feature matching for decision making, accounting for moderate geometrical transformations and possible lighting variations that often occur in image acquisition. See Fig. 7.3 for the approach framework. The proposed approach is extensively evaluated on NCUT Part A and NCUT Part B, both of which are among the largest dorsal hand vein datasets so far known in the literature. Experimental results clearly demonstrate the effectiveness of the proposed method.

The contributions of this study can be summarized as:

1. We prove that dorsal hand vein based people identification can not only rely on the vascular network, but also depend on the optical characteristics, *i.e.* the absorption and scattering properties, of surrounding skin as well as subcutaneous tissue, since the randomly inhomogeneous distribution of blood and various chromophores and pigments is subject dependent.

2. We interpret NIR dorsal hand images as landscapes and surfaces and identify these keypoints of their optical properties using geometric features through the quantities of multi-order (*i.e.* the 1st-, 2nd-, and 3rd-order) gradient cues, namely Harris cornerness measurement, Hessian blobness measurement, and curvature extrema by operating the DoG detector on a human vision inspired image representation, OGMs, which are closely related to the quantities in differential geometry.

3. We demonstrate that these keypoints as localized by the aforementioned multi-order gradient based quantities capture different geometric attributes corresponding to complementary facets of the optical properties of the vein network as well as its surrounding skin and subcutaneous tissue. As a result, we further propose to combine these local features for identification and achieve the best recognition accuracy so far known on the NCUT Part A dataset.

Following the processing steps in the framework, we first present the multi-level keypoint detection method, then describe the Oriented Gradient Maps (OGMs)
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.3: Framework of the proposed approach, including comprehensive representation of optical properties through multi-order gradient quantities and robust matching with SIFT-based features.

based dorsal hand representation, and after that we introduce the local matching step. The experimental results of both scenarios in dorsal hand vein recognition and verification are displayed and analysed finally.

7.4 Multi-level Keypoint Detection

For local feature-based matching approaches, keypoint detection is a critical step which is expected to locate a sufficient number of local feature points for a comprehensive description of the target image while providing some properties of invariance, e.g., scale, translation, rotation, etc. There exist several state of the art keypoint detection methods, such as Difference of Gaussian (DoG), Harris, Hessian, Harris-Laplace, Hessian-Laplace, whose properties on textured gray level images have been explicitly investigated by Roth and Winter [Roth & Winter 2008] in object retrieval (see Table 7.1). However, due to the optical properties, the dorsal hand vein image contains very few texture details. In this section, we are interested in the geometric attributes of these keypoint detectors when the underlying hand vein images are interpreted as surfaces. This geometric analysis results in the design of our multi-level keypoint detection for hand vein images.

7.4.1 The Difference of Gaussian Detector

Difference of Gaussian (DoG), proposed by Lowe [Lowe 2004], is one of the most widely used detectors, and it serves the SIFT feature extraction and matching.

The image is first repeatedly convolved with Gaussian filters of different scales separated by a constant factor, $k$, to generate an octave in the scale space. As for an input image, $I(x, y)$, its scale space is defined as a function, $L(x, y, \alpha)$, produced
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

by a convolution of a variable scale Gaussian $G(x, y, \alpha)$ with the input image $I$, and the DoG function $D(x, y, \alpha)$ can be computed from the difference of two nearby scales:

$$D(x, y, \alpha) = (G(x, y, k\alpha) - G(x, y, \alpha)) \times I(x, y)$$
$$= L(x, y, k\alpha) - L(x, y, \alpha)$$ \hspace{1cm} (7.1)

The extrema of $D(x, y, \alpha)$ can be detected by comparing each pixel value with those of its 26 neighbors within a $3 \times 3$ area at the current and adjacent scales. At each scale, gradient magnitude and orientation, $m(x, y)$ and $\theta(x, y)$ (as shown in (7.2) and (7.3)), are computed by exploiting pixel differences. The confirmed stable extremes are regarded as the scale-invariant keypoints located by DoG.

$$m^2(x, y) = (L(x + 1, y) - L(x - 1, y))^2$$
$$+ (L(x, y + 1) - L(x, y - 1))^2$$ \hspace{1cm} (7.2)

$$\theta(x, y) = \tan^{-1} \frac{L(x + 1, y) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)}$$ \hspace{1cm} (7.3)

DoG has proved competent at blob detection on gray level images. In hand vein analysis, the optical properties of the vein and its nearby tissue result in images with very limited texture details. Therefore, DoG locates very few feature points which are not located on hand vein regions (see Fig.7.4(a)). However, when the dorsal hand image is considered as a surface, we can give a geometric interpretation of these local features detected by using DoG. In our implementation, DoG can be regarded as an approximation of the Laplacian of Gaussian (LoG) with the ratio of the scales equal to 1.6. In this case, the Laplacian calculates the addition of these second partial derivatives and delivers the sum of both the curvatures in the $x$ and $y$ direction. Keeping this property in mind, we can see from Fig.7.4(a) that DoG has actually located a few points on the surface displayed through the dorsal hand image whose sums of the curvatures in the $x$ and $y$ directions are extrema either on ridges or on basins. Of course, the number of these feature points is not sufficient to comprehensively capture the whole geometric attributes of the underlying surface. We therefore study other state of the art local feature detection techniques and analyze their geometric properties.
Table 7.1: Comparison of different local feature detectors.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Invariance</th>
<th>Repeat</th>
<th>KeyPt.Num.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris</td>
<td>Rotation</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Hessian</td>
<td>Rotation</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Harris-Laplace</td>
<td>Rotation &amp; Scale</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Hessian-Laplace</td>
<td>Rotation &amp; Scale</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>DoG</td>
<td>Rotation &amp; Scale</td>
<td>High</td>
<td>Medium</td>
</tr>
</tbody>
</table>

### 7.4.2 Harris and Hessian Keypoint Detection

From Table 7.1, we can see that comparing with DoG, the other detectors, i.e. Harris, Hessian, Harris-Laplace, and Hessian-Laplace, not only have high repeatability, but also locate more keypoints on the gray level image. Both the Harris-Laplace and Hessian-Laplace are similar with DoG in the performance of scale invariance. Instead of Laplacian approximation, Harris- and Hessian-Laplace apply the scale normalized Laplacian to create the scale space which gives the benefit to local feature extraction and matching.

Specifically, the **Harris** detector was proposed by Harris and Stephens who defined the product of two first derivation matrices [Roth & Winter 2008]:

\[
M_{Ha} = \sigma_D^2 L^2 G(\sigma_I) \begin{bmatrix} L^2_x(p, \sigma_D) & L_xL_y(p, \sigma_D) \\ L_xL_y(p, \sigma_D) & L^2_y(p, \sigma_D) \end{bmatrix}
\]  

and it responds to corner features on gray level images. \( L_x \) and \( L_y \) denote the first derivation of the smoothed image \( L \) by convolving the raw image with Gaussian filter, whose standard deviation is \( \sigma_D \), at position \( p \) in the \( x \) and \( y \) direction respectively. The corner response threshold \( R_{Ha} \) calculated by avoiding the eigenvalue decomposition of the second moment matrix above by

\[
R_{Ha} = Det(M_{Ha}) - k \times Tr(M_{Ha})^2
\]  

The **Hessian** matrix-based detector is similar with the Harris detector but presents strong responses on blob features, instead of the corner ones, because the Hessian matrix-based detector replaces the elements of (7.4) with the second derivation:

\[
M_{He} = \sigma_D^2 L^2 G(\sigma_I) \begin{bmatrix} L_{xx}(p, \sigma_D) & L_{xy}(p, \sigma_D) \\ L_{xy}(p, \sigma_D) & L_{yy}(p, \sigma_D) \end{bmatrix}
\]  

138
where $L_{xx}$ and $L_{yy}$ are the second derivatives of the smoothed image $L$ at the position $p$ in the $x$ and $y$ direction respectively; and $L_{xy}$ is the mixed derivative in both directions.

The two detectors above, \textit{i.e.} Harris and Hessian, only own the property of rotation invariance. To achieve scale invariance, the scale normalized Laplacian $R_{\sigma}$ defined in (7.7) is introduced as a scale selection criterion by Harris-Laplace and Hessian-Laplace, and both detectors thus possess the property of scale invariant as DoG does.

$$R_{\sigma} = \sigma^2 \times |L_{xx}(p) + L_{yy}(p)| \quad (7.7)$$

While Harris and Hessian keypoint detectors are mostly analyzed in terms of properties of cornerness and blobness on gray level images, they can also be described as geometrical attributes on surfaces because of their close relationships with the first and second fundamental forms in differential geometry [Kuhnel 2005]. Indeed, the matrix of the Harris detector as defined in (7.4) is related to the symmetric matrix of the fundamental form which characterizes the metric properties of a surface, \textit{i.e.}, how the length and area are changed on the surface with regard to the ambient space. In other words, the matrix of Harris in (7.4) and the cornerness response in (7.5) characterize somehow the elasticity of a surface as we can see in Fig. 7.4(b). In this figure, the Harris-Laplacian detector locates much more keypoints in comparison with DoG. Furthermore, these keypoints cover the whole hand vein image and identify those points on the hand-dorsa surface whose elasticity is greater than a given threshold. They can thereby contribute to people identification using the optical properties of the hand dorsa subcutaneous tissue.

In regard to the Hessian-Laplace detector, the matrix in (7.6) with the second derivatives is related to the matrix of the second fundamental form which characterizes how an embedded surface is curved in the ambient space using curvature metrics, \textit{e.g.}, principal curvatures, mean and Gaussian curvatures. The Hessian-Laplace detector hence delivers keypoints on the hand dorsa surface with shape clues in terms of curvatures. As we can see in Fig. 7.4(c), the keypoints densely populate the valley regions, \textit{i.e.}, the hand vein regions.

### 7.4.3 Design of Multi-level Keypoint Detection

Given the fact that both the vein and the optical attributes of the surrounding subcutaneous tissue are subject dependent, an effective way to characterize a person is to adopt the keypoints localized by the Harris-Laplace and Hessian-Laplace detectors. The former captures the elasticity of the underlying surface of the dorsal hand whereas the latter delivers the points of shape information, in particular those
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.4: The distribution of keypoints detected by (a) DoG, (b) Harris-Laplace; and (c) Hessian-Laplace (1,000 clusters), on a hand-dorsa surface. DoG locates very few feature points whose sums of $x$ and $y$ curvatures are extrema; Harris-Laplace identifies the keypoints whose elasticities are greater than a threshold; Hessian-Laplace detects the keypoints which carry shape information in terms of curvatures, and localizes in particular the ones which densely populate the valley regions corresponding to veins.

Some statistical analysis has been conducted along with the experiments in this study using the images on the NCUT Part A database. DoG only detects less than 10 keypoints on each hand-dorsa image, and such a sparsity in local features cannot provide sufficient distinctiveness and thus fails to result in a reasonable recognition accuracy. When exploiting the Harris-Laplace detector, we can averagely locate 640 keypoints, and this number is indeed much larger than that of DoG.

Regarding Hessian-Laplace, around 3,000 local features can be found on each dorsal hand image, and this amount causes a sharp increase of computational cost in matching. We hence consider selecting a subset of the most representative features. Generally, it is straightforward to choose the strongest points in terms of their responses in detection. However, those points, achieved in this operation whose responses are higher than the others, only distribute on the partial dorsal hand vein network (as in Fig.7.5 (b)), leading to a loss in discriminative power. As a result, to reduce the number of Hessian-Laplace based points while keeping the distinctiveness, the number of keypoints is reduced through clustering (as shown in Fig.7.5 (a)), where only their locations (i.e. $x$ and $y$ coordinates) are considered. In our case, the k-means algorithm is employed to randomly cluster the points into 500, 700, and 1,000 respectively to balance the performance and efficiency.

140
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.5: 1000 selected keypoints located by the Hessian-Laplace detector: (a) based on clustering, and (b) based on strongest responses.

7.5 Oriented Gradient Maps based Representation

In the previous section, we analyze the geometric properties of several state of the art keypoint detectors when dorsal hand vein images are interpreted as surfaces, and propose a multi-level keypoint detector to localize features not only on the vein but also on its surrounding subcutaneous tissue. In this section, we further increase the descriptive completeness of these local features through an approach inspired by human vision, using Oriented Gradient Maps (OGMs) which are originally applied to represent the texture as well as shape information for 3D face recognition [Huang et al. 2011a].

The objective of the OGMs is to provide a visual description simulating the operation of human complex cells in the visual cortex [Hubel & Wiesel 1962a]. These complex neurons respond to a gradient at a particular orientation and spatial frequency, but the location of the gradient is allowed to shift over a small receptive field rather than being precisely localized.

7.5.1 Representation of Complex Neuron Response

The proposed OGM based representation simulates the response of complex neurons through a convolution of gradients in specific directions within a pre-defined neighborhood. Since the scale of the dorsal hand vein image changes slightly thanks to the hardware setup, we only employ a circular neighborhood $R$, as demonstrated in Fig. 7.6. The precise radius value of the circular area needs to be fixed experimentally. The response of a complex neuron at a given pixel location is a set of gradient maps in different orientations convolved by a Gaussian kernel.

Specifically, given an input image (a dorsal hand vein image in our case) $I$, a certain number of gradient maps $G_1, G_2, ..., G_o$, one for each quantized direction $o$, are firstly computed. They are defined as:
The neighborhood of the complex neurons is a circular area and its radius can be changed according to the scale.

\[ G_o = \left( \frac{\partial I}{\partial o} \right)^+ \quad (7.8) \]

The “+” sign indicates that only the positive values are kept to preserve the polarity of the intensity changes, and the negative ones are set to zero.

Each of gradient maps describes gradient norms of the input image in an orientation \( o \) at every pixel. We further simulate the response of the complex neurons by convolving its gradient maps with a Gaussian kernel \( G \), and its standard deviation is proportional to the radius value of the given neighborhood, \( R \), as in (7.9).

\[ \rho_o^R = G_R * G_o \quad (7.9) \]

The purpose of the convolution with Gaussian kernels is to allow the gradients to shift in a neighborhood without abrupt changes.

At a given pixel location \((x, y)\), we collect all values of the convolved gradient maps at that location and form the vector \( \rho^R(x, y) \), and it hence possesses a response value of complex neurons for each orientation \( o \).

\[ \rho^R(x, y) = [\rho_1^R(x, y), \ldots, \rho_o^R(x, y)]^t \quad (7.10) \]

\( \rho^R(x, y) \), is then normalized to an unit norm vector, which is called response vector and denoted by \( \rho^R \).
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.7: The OGMs describe a perceived Near-Infrared hand-dorsa vein image in 8 orientations.

7.5.2 Oriented Gradient Maps by Response Vectors

According to the definition of the response vector, the dorsal hand vein image can be represented by its perceived values of complex neurons. Specifically, given a hand vein image \( I \), we generate an Oriented Gradient Map (OGM) \( J_o \) using complex neurons for each orientation \( o \) defined as in (7.11).

\[
J_o(x, y) = \rho^R_\omega(x, y)
\]

Fig.7.7 depicts such a process applied to a dorsal hand vein image. In this study, we generate eight OGMs for eight pre-defined quantized directions. Instead of the original NIR hand dorsa images, these OGMs are thus exploited in the subsequent local feature extraction and matching for identification.

7.5.3 The Properties of Distinctiveness and Invariance

The OGMs potentially offer high distinctiveness since they highlight the details of local texture changes. Meanwhile, they also possess some interesting properties of robustness to affine lighting variations.

When applied OGMs to dorsal hand vein images, they offer the property of being robust to affine illumination transformations. Indeed, each OGM, \( J_o \), is simply normalized convolved gradient maps at the orientation \( o \), while monotonic
illumination change often adds a constant intensity value, as a result, it does not affect the computation of gradients. Furthermore, a change in image contrast in which the intensities of all pixels are multiplied by a constant will lead to the multiplication of gradient computation; however, such a contrast change will be cancelled by the normalization of response vectors.

OGMs can be made even rotation invariant if we choose to quantize directions starting from that of the principal gradient of all the gradients within the neighborhood, and the tolerance to scale variations can also be largely improved by embedding the multi-scale strategy. Nevertheless, we do not perform such steps to save computational cost as the dorsal hand vein images in our study were already roughly aligned.

7.5.4 Design of the OGM based Keypoint Detector

After the OGMs of a dorsal hand vein image are computed to highlight the details of optical properties of the underlying vein network and its nearby subcutaneous tissue, they are then interpreted as retinal images [Morgan 2011], i.e., surfaces or landscapes, and their geometric attributes can be further analyzed. In this work, we concentrate on the variations of shape of these OGM-based surfaces and employ DoG which identifies the keypoints whose sums of curvatures in the $x$ and $y$ directions change the most.

Fig.7.8 demonstrates the distribution of the keypoints detected by DoG, from the hand-dorsa vein image and its corresponding OGMs, respectively. Because OGMs simulate the operation of complex cells of the visual cortex and therefore highlight the details of the vein patterns and their surrounding subcutaneous tissue, DoG locates much more keypoints, including in particular the dorsal hand vein minutia, on these OGM-based surfaces in comparison with the smooth raw hand-dorsa vein surface, which comprehensively describe these optical characteristics. The statistics that we computed show that the average number of keypoints extracted from each of OGM can rise up to 627, while that from the original dorsal hand vein image is less than 10 as stated in Section III. Fig.7.8 illustrates this phenomenon.

Recall that the Harris-Laplace detector locates the keypoints whose elasticities are greater than a threshold from the hand-dorsa surface and the Hessian-Laplace detector mostly localizes the keypoints in the valley regions, i.e., the vein areas. As compared to these features located by the multi-level detector, i.e., Harris-Laplace as well as Hessian-Laplace, DoG identifies the keypoints whose shape changes the most at a given OGM-based surface. In the viewpoint of differential calculus, Hessian-Laplace provides the first order gradient information; Hessian-Laplace offers the second order gradient information; whereas DoG associated with OGMs generates
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.8: Comparison in keypoint detection by DoG in the raw hand-dorsa vein image (center) and its corresponding OGMs in the eight pre-defined quantized orientations (around).

the third order gradient information from an input hand-dorsa surface. These keypoints are thus complementary for a comprehensive description of the hand-dorsa vein image and can be used for people identification through a local feature matching process.

7.6 Local Feature Matching

Once identified the keypoints using the multi-level keypoint detection and DoG with OGM as described in the previous two sections, we further extract the widely-used SIFT features [Lowe 2004] at these positions to enable the matching between two dorsal hand vein images for similarity measure computation and final decision making.

7.6.1 SIFT-feature based Matching

For each detected keypoint, a feature vector is extracted as a descriptor from these gradients of sampling points within its neighborhood. In order to obtain the orientation invariance, the coordinates and gradient orientations of sampling points in the neighborhood are rotated relative with the keypoint orientation. Then a Gaussian function is employed to assign a weight to the gradient magnitude of each point.
Points close to the keypoint are given more emphasis than the ones far from it (see [Lowe 2004] for more details about the SIFT parameter setting). The orientation histograms of $4 \times 4$ sampling regions are calculated, each with eight orientation bins. Thus a feature vector with a dimension of 128 ($4 \times 4 \times 8$) is produced.

Given these local features extracted from the original image pair or each of their corresponding OGM pairs in the gallery and probe sets respectively, the two sets of keypoints on dorsal hands can be associated. Matching one keypoint to another is accepted only if the similarity distance is below a pre-defined threshold $t$ times the distance to the second closest match. In this work, $t$ is empirically set at 0.6 as in [Lowe 2004]. The number of matched keypoints is accounted as the similarity measurement between the gallery and probe samples, and a larger matching score indicates a bigger probability that the hand-dorsa images are from the same hand.

### 7.6.2 Score Level Fusion

For a hand-dorsa vein image, we extract a set of keypoints of multiple sources, i.e., the multi-level detection based ones directly localized on the original image by the Harris-Laplace and Hessian-Laplace detectors as well as the ones detected by DoG on its corresponding OGMs at different orientations. As a result, multi-source keypoint matches can be associated for identification. We then combine their similarity measurements at the matching score level to take all these contributions into account for final decision making.

Specifically, we denote the number of the matched keypoints by $N_{Harr}$ for the ones localized using Harris-Laplace and by $N_{Hess}$ for the ones detected employing Hessian-Laplace from the original hand-dorsa vein image pair; and by $N_{OGM_o}$ for the ones found exploiting DoG from each of their corresponding OGM pairs at the $o_{th}$ direction. The bigger the value of $N$ is, the more likely that the two dorsal hand images belong to the same subject, indicating that the similarity measurements, i.e. $N_{Harr}$, $N_{Hess}$, and $N_{OGM_o}$, are all with the positive polarity (a bigger value means a better matching relationship). A dorsal hand vein image in the probe set is compared with the ones in the gallery set respectively, leading to a matching score vector. The $n_{th}$ element in a matching score vector corresponds to the similarity between the probe and the $n_{th}$ gallery sample. The score vectors from multiple sources are further normalized to the interval of $[0, 1]$ using the max-min rule. These matching scores are finally fused by a basic weighted sum rule:

$$S = \sum_{i=1}^{o+2} w_i \cdot S_i \quad \text{(7.12)}$$

There are totally $o + 2$ similarity scores including the one of $S_{Harr}$, the one
of $S_{Hess}$, and the ones ($o$) of $S_{OGM_o}$. The corresponding weight $w_i$ is calculated dynamically during the online step using the scheme as in [Mian et al. 2008]:

$$w_i = \frac{\max_1(S_i) - \text{mean}(S_i)}{\max_2(S_i) - \text{mean}(S_i)}$$  \hspace{1cm} (7.13)

where the operators $\max_1(S)$ and $\max_2(S)$ produce the first and second maximum values of the score $S$ respectively. The gallery dorsal hand vein image that holds the maximum value is declared as the identity of the probe image.

### 7.6.3 Illustration of Matching Samples

Fig.7.9 displays a matching example adopting multi-level keypoint detection applied directly to the dorsal hand vein image: i.e. Harris-Laplace (Fig.7.9(a)) and Hessian-Laplace (Fig.7.9(b)). According to the locations of the matched keypoints, they are highlighted by using two different colors. The ones in the vein area are marked in yellow whilst the ones in the surrounding subcutaneous tissue are marked in red. This figure highlights the following facts: (1) the positions of the keypoints provided by both the detectors, i.e. Harris-Laplace and Hessian-Laplace, are different and the two sets of points are complementary to each other; and (2) these details outside the vein regions, i.e., these matched keypoints in the red boxes, are as important as those within the vein regions in final decision making.

Fig.7.10 depicts a matching example between two hand-dorsa images of the same subject utilizing OGMs with DoG. We can observe a similar phenomenon and draw the same conclusion, i.e., the clues conveyed within the vein region (marked in yellow) and its surrounding subcutaneous tissue (marked in red) are both discriminative; meanwhile, these OGMs at different orientations contain complementary information.

### 7.7 Experiments

In order to comprehensively evaluate the proposed method, we designed several experiments that are explicitly introduced in the subsequent. The experiments (in the Subsection A to F) were mainly conducted both in the scenarios of identification and verification as in the state-of-the-art work using the NCUT Part A database. In the meantime, to check the generalization ability of the proposed approach, we also carried out additional experiments (in the Subsection G) on NCUT Part B collected by a device whose cost is only a half as that for Part A, and its images are thus with more noise. Recall that both databases are among the largest ones of NIR hand-dorsa vein images. Part A contains 10 right and 10 left dorsal hand
Figure 7.9: A matching example between the dorsal hand vein images belonging to the same person based on these keypoints detected using (a) Harris-Laplace and (b) Hessian-Laplace. The matched keypoints marked in yellow boxes are located in the vein region and the ones in red boxes are located in the nearby subcutaneous tissue.

Figure 7.10: A matching example using these OGM pairs of two left hands of the same person. The left column from top to bottom: OGM1 to OGM4; while the right column with the same order: OGM5 to OGM8. The matched keypoints marked in yellow are located in the vein area and the ones marked in red are detected in the nearby subcutaneous tissue.
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Table 7.2: The results of different detectors, i.e. Harris-Laplace and Hessian-Laplace and their different fusion schemes for the multi-level keypoint detection based method on NCUT Part A.

<table>
<thead>
<tr>
<th>Harris</th>
<th>Hessian</th>
<th>Sum</th>
<th>Product</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>500kpt.</td>
<td>97.43%</td>
<td>87.65%</td>
<td>95.49%</td>
<td>86.08%</td>
<td></td>
</tr>
<tr>
<td>640kpt.</td>
<td>97.55%</td>
<td>97.55%</td>
<td>95.49%</td>
<td>94.61%</td>
<td></td>
</tr>
<tr>
<td>700kpt.</td>
<td>93.63%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000kpt</td>
<td>97.55%</td>
<td>98.04%</td>
<td>95.88%</td>
<td>96.18%</td>
<td></td>
</tr>
</tbody>
</table>

images respectively for each of the 102 subjects (totally 2,040 samples), while Part B consists of the same number of images from both hands of 101 subjects (totally 2,020 samples). All the hand images were roughly aligned thanks to the hardware configuration, but they still have moderate viewpoint (i.e. rotation and translation) and slight lighting intensity variations.

7.7.1 Effectiveness of Multi-Level Keypoint Detection

We evaluated the effectiveness of the proposed multi-level keypoint detection approach in terms of the rank-one recognition rate in the scenario of identification. For experimental setup, the first five images of a subject were used in the gallery set and the remaining five images were exploited as probes. Because it was found out that the hand vein pattern is unique to some level for each person and each hand [Badawi 2006], we considered the left and right hand-dorsa vein images separately as if we had 204 different subjects each of which possesses 10 samples in the dataset.

From the results in Table 7.2, we can conclude in these points:

- When we increase the number of these clustered centers (i.e. from 500 to 1000) for the keypoints detected by Hessian-Laplace, the rank-one recognition rate is improved, indicating that more keypoints lead to better accuracy. As we continue to increase it, the improvement is more and more limited. We thus set this number at 1,000 to balance the accuracy and time cost in the following experiments to compute the performance of Hessian-Laplace.

- Making use of a comparable number of detected keypoints, the performance achieved by Harris-Laplace is superior to that of Hessian-Laplace, demonstrating that the keypoints localized by Harris-Laplace, including in particular the ones outside the vein areas, provide more discriminative information than those detected by Hessian-Laplace which mainly focuses on the vein regions. This phenomenon further illustrates the fact that these optical properties of subcutaneous tissue surrounding the vein network convey subject dependent cues.

- No matter which score level fusion scheme (sum, product, max, and min rule)
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Table 7.3: Performance of each OGM and their combination in the setup of left-hand only, right-hand only and both-hands on the NCUT Part A database.

<table>
<thead>
<tr>
<th>Directions</th>
<th>Left Hand</th>
<th>Right Hand</th>
<th>Both Hands</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGM-1</td>
<td>92.94%</td>
<td>93.53%</td>
<td>93.04%</td>
</tr>
<tr>
<td>OGM-2</td>
<td>81.18%</td>
<td>79.41%</td>
<td>78.53%</td>
</tr>
<tr>
<td>OGM-3</td>
<td>75.88%</td>
<td>77.45%</td>
<td>75.10%</td>
</tr>
<tr>
<td>OGM-4</td>
<td>73.14%</td>
<td>60.78%</td>
<td>66.18%</td>
</tr>
<tr>
<td>OGM-5</td>
<td>97.57%</td>
<td>92.55%</td>
<td>91.57%</td>
</tr>
<tr>
<td>OGM-6</td>
<td>78.82%</td>
<td>74.90%</td>
<td>75.49%</td>
</tr>
<tr>
<td>OGM-7</td>
<td>77.65%</td>
<td>84.51%</td>
<td>80.69%</td>
</tr>
<tr>
<td>OGM-8</td>
<td>78.63%</td>
<td>82.16%</td>
<td>78.82%</td>
</tr>
<tr>
<td>Fusion</td>
<td>99.02%</td>
<td>99.02%</td>
<td>99.02%</td>
</tr>
</tbody>
</table>

we take, the recognition rate is better than either of the Harris-Laplace or Hessian-Laplace, proving that the two detectors provide complementary clues to each other and highlighting the effectiveness of the multi-level keypoint detection approach. To keep the consistency in our approach, the sum rule was used in the following experiments to combine the results of Harris- and Hessian-Laplace.

7.7.2 Discriminative Power of OGMs

We then tested the discriminative power of the OGM based image representation in terms of the rank-one recognition rate in the identification scenario as well, following the same protocol as in the previous experiment. We calculated recognition rates of each OGM (for different quantized orientations) and their combination as displayed in Table 7.3.

As we discussed in section IV, each dorsal hand image has quite limited number of keypoints if DoG is directly applied to the original data, thus leading to a very partial description for the following matching step. This observation was our major motivation to develop OGMs which simulate the response of complex cells in the visual cortex in highlighting the gradients at different orientations. In Table 7.3 we can see that the fusion of all these OGMs reaches a much better result than any of the single one. Such a fact accords with our preliminary study for this issue in adopting subspace techniques [Chen et al. 2011]. Unfortunately, in that work, due to the sensitivity of holistic methods to NIR intensity variations and hand geometric transformations, only about 70%-80% rank-one recognition rates were reported even with an easier experimental setup. Obviously, that performance is not accurate enough for a biometric system.

Meanwhile, we can see that the results of these OGMs are different, and the ones of OGM-1 and 5 are largely better than the others. The reason lies in that most of
the dorsal hand veins are vertically distributed as shown in Fig. 7.8, which can be best highlighted by the horizontal gradient responses, i.e. OGM-1 and 5. Moreover, there exist a few horizontal and oblique vein furtations, and their corresponding best gradient responses are also necessary to comprehensively represent the entire venous network. As a result, the joint use of all these OGMs leads to the final highest score, indicated by the fusion performance.

On the other hand, we compared these results in the three columns of Table 7.3, and found out that the performance only using left hand images was comparable to that only using right hand images. When left and right hand vein images were both used and considered as captured from different subjects, the result generally remains stable, showing that our method works well as the class size is doubled.

### 7.7.3 Impact of Gallery Size

An important property of a biometric system is its stability when the gallery size changes. For this purpose, we varied the number of gallery samples of each person from 1 to 9 (since at least 1 sample per person should be used in the probe set) to analyze the impact of the gallery size on the proposed method, employing the pre-defined experimental setup in identification. We can find that the rank-one recognition rate based on multi-level keypoint detection decreases from 97.55% to 85.57% (as in Fig. 7.11) and the accuracy by combining these OGMs at all orientations falls from 99.02% to 83.88% (as in Fig. 7.12) when the gallery size drops from 5 to 1. It indicates that the problem of limited enrolled samples seriously challenges the biometric system. In the meantime, we can also see that they both display certain robustness to such a challenge, and achieve acceptable rank-one recognition rates at 92.52% and 91.42% respectively, when only 2 samples were enrolled as gallery for each subject.

We highlighted in the previous sections the complementarity of the two solutions proposed in this study, i.e. the detection of multi-level keypoints which focuses on the elasticity and shape changes on the original hand-dorsa surface and DoG applied to its OGMs which simulate the response of complex cells in the visual cortex in highlighting these details through gradients at different orientations. A natural alternative to further improve the accuracy of the approach is to combine these scores of the two solutions to account for both their descriptive power. We adopt the weighted sum rule as defined in (7.12) for fusion, and final performance is significantly improved. The entire system reports a rank-one recognition rate up to 91.29% (as shown in Fig. 7.13 and Table 7.4) with only one image of each subject in the gallery set. As we can see in Table. 7.4 and Fig. 7.13, these multi-source keypoints are consistently complementary.

These cumulative match characteristic (CMC) curves related to different numbers (from 1 to 9) of enrolled samples in the gallery set of each subject are provided
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.11: Accuracy curves based on multi-level keypoint detection with respect to the gallery size of each subject on NCUT Part A.

Figure 7.12: Accuracy curves based on individual OGMs and their combination with respect to the gallery size of each subject on NCUT Part A.

7.7.4 Verification Validation

We also performed experiments in the scenario of verification with the three modalities, i.e. multi-level keypoint detection on the original images, DoG based keypoint detection on these OGMs of the images, as well as the combination of the previous two modalities. For each subject, the first image was regarded as the gallery and the remaining images were treated as the probe samples to calculate the Verification Rates (VR) at the False Acceptance Rate (FAR) of 0.001 and the Equal Error Rate (EER). Table 7.5, Fig. 7.15 and 7.16 display these results, from which we can draw similar conclusions as in identification.
Chapter 7. Hand-dorsa Vein Recognition by Matching Local Features of Multi-source Keypoints

Figure 7.13: Accuracy curves based on different sources of keypoint matching with respect to the gallery size of each subject on NCUT Part A.

Figure 7.14: CMC curves based on multi-source keypoint matching of different numbers of gallery samples of each subject on NCUT Part A.

Table 7.4: The rank-one recognition rates of keypoint matching in each source as well as their combination, i.e. multiple sources, with respect to the gallery size of each subject on NCUT Part A.

<table>
<thead>
<tr>
<th>GalleryNum.</th>
<th>Harris</th>
<th>Hessian</th>
<th>OGMs</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77.67%</td>
<td>77.89%</td>
<td>83.88%</td>
<td>91.29%</td>
</tr>
<tr>
<td>2</td>
<td>86.95%</td>
<td>88.42%</td>
<td>91.42%</td>
<td>96.38%</td>
</tr>
<tr>
<td>3</td>
<td>90.83%</td>
<td>91.74%</td>
<td>94.82%</td>
<td>98.18%</td>
</tr>
<tr>
<td>4</td>
<td>94.69%</td>
<td>94.53%</td>
<td>97.88%</td>
<td>99.35%</td>
</tr>
<tr>
<td>5</td>
<td>96.27%</td>
<td>95.59%</td>
<td>99.02%</td>
<td>99.61%</td>
</tr>
<tr>
<td>6</td>
<td>97.92%</td>
<td>96.45%</td>
<td>99.14%</td>
<td>99.39%</td>
</tr>
<tr>
<td>7</td>
<td>98.86%</td>
<td>97.06%</td>
<td>99.67%</td>
<td>99.84%</td>
</tr>
<tr>
<td>8</td>
<td>98.77%</td>
<td>97.30%</td>
<td>99.75%</td>
<td>100.00%</td>
</tr>
<tr>
<td>9</td>
<td>99.02%</td>
<td>98.04%</td>
<td>99.51%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Figure 7.15: ROCs between FAR and VR of the proposed method on the NCUT Part A dataset.

Figure 7.16: ROCs between FAR and FRR of the proposed method on the NCUT Part A dataset.

Table 7.5: Performance in the scenario of verification of the proposed method on the NCUT Part A dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>FAR</th>
<th>VR</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-level Keypoint</td>
<td>0.001</td>
<td>91.99%</td>
<td>1.91%</td>
</tr>
<tr>
<td>OGM based keypoints</td>
<td>0.001</td>
<td>93.63%</td>
<td>1.36%</td>
</tr>
<tr>
<td>Multi-source Keypoints</td>
<td>0.001</td>
<td>96.35%</td>
<td>0.81%</td>
</tr>
</tbody>
</table>
Table 7.6: Comparison with the state of the art in rank-one recognition rate on the NCUT Part A dataset.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Class</th>
<th>Gall./Prob.</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-source Keypoint+SIFT</td>
<td>204</td>
<td>816/1224</td>
<td>99.35%</td>
</tr>
<tr>
<td>Binary+SIFT [Wang et al. 2012]</td>
<td>204</td>
<td>816/1224</td>
<td>78.68%</td>
</tr>
<tr>
<td>Best Binary+SIFT [Wang et al. 2012]</td>
<td>204</td>
<td>816/1224</td>
<td>97.95%</td>
</tr>
<tr>
<td>Multi-source Keypoint+SIFT</td>
<td>204</td>
<td>1020/1020</td>
<td>99.61%</td>
</tr>
<tr>
<td>Multi-level Keypoint + SIFT [Tang et al. 2012]</td>
<td>204</td>
<td>1020/1020</td>
<td>98.04%</td>
</tr>
<tr>
<td>OGMs + SIFT [Huang et al. 2012a]</td>
<td>204</td>
<td>1020/2020</td>
<td>99.02%</td>
</tr>
<tr>
<td>LCP [Wang et al. 2010a]</td>
<td>204</td>
<td>1020/1020</td>
<td>90.88%</td>
</tr>
<tr>
<td>LBP+Graph [Zhu &amp; Huang 2012]</td>
<td>204</td>
<td>1020/1020</td>
<td>96.67%</td>
</tr>
<tr>
<td>MLBP [Zhao et al. 2008]</td>
<td>15</td>
<td>150</td>
<td>97.30%</td>
</tr>
</tbody>
</table>

7.7.5 Comparison with the State of the Art

We compared the proposed method with the state of the art ones on NCUT Part A as illustrated in Table 7.6. Specifically, in [Wang et al. 2012], Wang et al. firstly localized the vein network on the dorsal hand; then represented the detection result as a binary image, and finally applied SIFT for the matching step. Such an approach was originally introduced by Ladoux et al. [Ladoux et al. 2009] for the purpose of hand-palm vein identification. As we can see from that table, when only the vein regions are used as in [Wang et al. 2012], the accuracy is only 78.68% with the first 4 samples in the gallery and the other 6 as probes, thereby far behind the performance achieved by the proposed approach. This comparison confirms once more the importance of considering the optical properties of the whole hand-dorsa image. Our result is also higher than the best one reported in [Wang et al. 2012] that was achieved by adopting the relationship of multiple gallery samples of each subject. In [Wang et al. 2010a], Wang et al. employed the circular partition LBP, namely LCP, and achieved a recognition rate of 90.88% with 5 hand vein images in the gallery set and the remaining 5 ones used as probes. With this protocol, Zhu and Huang [Zhu & Huang 2012] evaluated their approach via hierarchically combining the LBP based texture features and graph matching based geometric features, and a rank-one recognition rate of 97.67% was reported. In our case, a better result is obtained by using such an experimental setup. A comparable performance was generated in [Zhao et al. 2008], but in their experiments only a subset of the dataset (150 gallery and probe images of 15 persons) was exploited for evaluation. These facts clearly demonstrate the effectiveness of the proposed approach for dorsal hand vein recognition.

For further information and comparison, Table 7.7 summarizes major state of the art approaches for the issue of dorsal hand vein based people identification and
verification. As we can see, the proposed method achieves competitive results in both the scenarios of identification and verification while using a more comprehensive dataset.

### 7.7.6 Complementarity of Left and Right Hands

Since vein patterns are different to some level for both hands of the same individual [Badawi 2006], intuitively, the left and right hands of one person should possess complementary information for recognition. In the experiment, we further investigated such an answer to this problem by fusing the similarity measurement of each hand using the weighted sum rule as the other fusion steps in this study. We can see from Table 7.8 that the accuracy based on the fusion of both hands (in the 3rd column) always outperforms that based on either of single hand (in the 1st and 2nd columns), and we can achieve a rank-one recognition rate of 98.15% even using one enrolled sample per hand for each subject. These results thus suggest that the use of left and right hands can further reinforce the robustness and the performance of the proposed hand vein based biometric system.

### 7.7.7 Evaluation on Generalization Ability

The previous experimental results suggest that the proposed method achieves very good performance on the NCUT Part A dataset. A key question, nevertheless, is whether it generalizes to other databases? We aim to answer this question using the novel NCUT Part B dataset. It roughly keeps the same size as NCUT Part A, but the hand vein images are more noisy since they were captured using a low cost device. We conducted both scenarios of identification as well as verification, and adopted the same protocols as in Table 7.4 and 7.5 respectively.

We can observe from Table 7.9 and 7.10 that the conclusions achieved on NCUT Part B are consistent with the ones on Part A, *i.e.* the joint utilization of multiple keypoints to characterize these optical attributes of entire dorsal hand images improves the performance in comparison with those local feature-based approaches using single ones. Furthermore, when we compare the accuracies on NCUT Part A and NCUT Part B (*i.e.* Table 7.4 vs. Table 7.9 and Table 7.5 vs. Table 7.10), it can be seen that the proposed method reports competitive accuracies on NCUT Part B as well, quite close to the ones achieved on NCUT Part A. These results thereby suggest a quite good generalization skill of the proposed approach to dorsal hand vein recognition which further displays the robustness to noise caused by the cost decrease in the device of data acquisition.
Table 7.7: Comparative summary of related work on dorsal hand vein based identification and verification on different databases.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Approach</th>
<th>Imaging</th>
<th>Database</th>
<th>Identification (G:P)</th>
<th>Verification (FAR; FRR; G:P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995 Cross and Smith</td>
<td>Sequential correlation on vein signatures</td>
<td>NIR</td>
<td>20 Hands (5 Images)</td>
<td>–</td>
<td>0.00%; 5.00%; (3:2)</td>
</tr>
<tr>
<td>2004 Tanaka and Kubo</td>
<td>FFT phase correlation and template matching</td>
<td>NIR</td>
<td>25 Hands (Unknown)</td>
<td>–</td>
<td>0.73%; 4.00%; (Unknown)</td>
</tr>
<tr>
<td>2004 Lin and Fan</td>
<td>Multi-resolution filter representation</td>
<td>FIR</td>
<td>32 Hands (20 Images)</td>
<td>–</td>
<td>2.30%; 2.30%; (5:15)</td>
</tr>
<tr>
<td>2005 Wang and Leedham</td>
<td>Hausdorff distance based line segment matching</td>
<td>FPR</td>
<td>12 Hands (9 Images)</td>
<td>–</td>
<td>0.00%; 0.00%; (3:6)</td>
</tr>
<tr>
<td>2005 Ding et al.</td>
<td>Distances between crossing and end points</td>
<td>NIR</td>
<td>48 Hands (5 Images)</td>
<td>–</td>
<td>0.00%; 0.90%; (1120 Matches)</td>
</tr>
<tr>
<td>2006 Wang et al.</td>
<td>Multi-supplemental features and multi-classifier fusion</td>
<td>NIR</td>
<td>100 Hands (5 Images)</td>
<td>–</td>
<td>0.00%; 0.50%; (3:2)</td>
</tr>
<tr>
<td>2008 Wang et al.</td>
<td>Modified Hausdorff distance on segmented vein area</td>
<td>NIR</td>
<td>47 Hands (3 Images)</td>
<td>–</td>
<td>0.00%; 0.00%; (Unknown)</td>
</tr>
<tr>
<td>2008 Wang et al.</td>
<td>Line Edge Mapping (LEM)/Gabor</td>
<td>NIR</td>
<td>100 Hands (5 Images)</td>
<td>66%/80%; (1:4)</td>
<td>–</td>
</tr>
<tr>
<td>2009 Kumar and Prathyusha</td>
<td>Shape feature and triangulation</td>
<td>NIR</td>
<td>100 Hands (3 Images)</td>
<td>–</td>
<td>1.14%; 1.14%; (2:1)</td>
</tr>
<tr>
<td>2010 Yüksel et al.</td>
<td>Independent Component Analysis (ICA)</td>
<td>NIR</td>
<td>100 Hands (3 Images)</td>
<td>94.16%; (1:2)</td>
<td>2.47%; 2.47%; (1:2)</td>
</tr>
<tr>
<td>2012 Hsu et al.</td>
<td>Block based 2D PCA and 2D LDA</td>
<td>NIR</td>
<td>214 Hands (20 Images)</td>
<td>98.55%; (10:10)</td>
<td>–</td>
</tr>
<tr>
<td>This paper</td>
<td>Multi-source keypoint based SIFT matching</td>
<td>NIR</td>
<td>204 Hands (10 Images)</td>
<td>99.61%; (5:5)</td>
<td>0.80%; 0.80%; (1:9)</td>
</tr>
</tbody>
</table>
Table 7.8: The results of left hand only, right hand only, and their fusion using different numbers of gallery samples on the NCUT Part A dataset.

<table>
<thead>
<tr>
<th>Gallery Num.</th>
<th>Left Hand</th>
<th>Right Hand</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91.72%</td>
<td>91.18%</td>
<td>98.15%</td>
</tr>
<tr>
<td>2</td>
<td>96.94%</td>
<td>95.59%</td>
<td>99.51%</td>
</tr>
<tr>
<td>3</td>
<td>98.46%</td>
<td>97.76%</td>
<td>100.00%</td>
</tr>
<tr>
<td>4</td>
<td>99.35%</td>
<td>99.51%</td>
<td>100.00%</td>
</tr>
<tr>
<td>5</td>
<td>99.61%</td>
<td>99.41%</td>
<td>100.00%</td>
</tr>
<tr>
<td>6</td>
<td>99.51%</td>
<td>99.26%</td>
<td>100.00%</td>
</tr>
<tr>
<td>7</td>
<td>99.67%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>8</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>9</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 7.9: The rank-one recognition rates of keypoint matching in each source as well as their combination, i.e. multiple sources, with respect to the gallery size of each subject on NCUT Part B.

<table>
<thead>
<tr>
<th>Gallery Num.</th>
<th>Harris</th>
<th>Hessian</th>
<th>OGMs</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.02%</td>
<td>73.82%</td>
<td>79.26%</td>
<td>89.77%</td>
</tr>
<tr>
<td>2</td>
<td>82.98%</td>
<td>82.98%</td>
<td>90.90%</td>
<td>94.06%</td>
</tr>
<tr>
<td>3</td>
<td>85.01%</td>
<td>86.14%</td>
<td>94.27%</td>
<td>95.76%</td>
</tr>
<tr>
<td>4</td>
<td>88.45%</td>
<td>87.95%</td>
<td>95.54%</td>
<td>96.78%</td>
</tr>
<tr>
<td>5</td>
<td>91.78%</td>
<td>90.20%</td>
<td>96.93%</td>
<td>97.82%</td>
</tr>
<tr>
<td>6</td>
<td>94.55%</td>
<td>91.46%</td>
<td>98.02%</td>
<td>98.14%</td>
</tr>
<tr>
<td>7</td>
<td>95.05%</td>
<td>93.07%</td>
<td>98.42%</td>
<td>98.51%</td>
</tr>
<tr>
<td>8</td>
<td>96.53%</td>
<td>93.81%</td>
<td>99.01%</td>
<td>99.13%</td>
</tr>
<tr>
<td>9</td>
<td>97.52%</td>
<td>95.54%</td>
<td>99.01%</td>
<td>99.50%</td>
</tr>
</tbody>
</table>

Table 7.10: Performance in the scenario of verification of the proposed method on the NCUT Part B dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>FAR</th>
<th>VR</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-level Keypoint</td>
<td>0.001</td>
<td>90.29%</td>
<td>3.21%</td>
</tr>
<tr>
<td>OGM based keypoints</td>
<td>0.001</td>
<td>91.38%</td>
<td>3.08%</td>
</tr>
<tr>
<td>Multi-source Keypoints</td>
<td>0.001</td>
<td>96.31%</td>
<td>1.12%</td>
</tr>
</tbody>
</table>
Table 7.11: Average consumed time of each component of the dorsal hand vein recognition system.

<table>
<thead>
<tr>
<th>System Component</th>
<th>Average Consumed Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keypoint Detection</td>
<td></td>
</tr>
<tr>
<td>Harris-Laplace</td>
<td>46ms</td>
</tr>
<tr>
<td>Hessian-Laplace</td>
<td>72ms</td>
</tr>
<tr>
<td>DoG on OGMs</td>
<td>66ms</td>
</tr>
<tr>
<td>SIFT Feature Extraction</td>
<td>52ms</td>
</tr>
<tr>
<td>Matching</td>
<td>4ms</td>
</tr>
</tbody>
</table>

### 7.7.8 Complexity Analysis

Real-life applications require a fast runtime in field deployment. The major cost centers in our system lie in multi-source keypoint detection, SIFT feature extraction, and matching. We therefore focused our attention on optimizing these procedures. Primarily, we optimally coded them using C++. Given the fact that each process on a single type of keypoints, i.e., the ones provided by Harris-Laplace, Hessian-Laplace, and OGMs, is separated, we then utilized multi-threading implementation of the entire system so that they can operate in parallel. In each process, the matching stage was further optimized and made 20 times faster in efficiency through GPU. Finally, the overall time cost is the maximum of the individual processes (i.e. the Hessian-Laplace based process). Table 7.11 shows the details.

As we can see in Table 7.11, the method currently costs about 128ms to achieve a 1-to-1 verification, including multi-source keypoint detection (72ms for the Hessian-Laplace based step) and SIFT feature extraction (52ms) on the given probe and its matching with the gallery (4ms), by using a machine equipped with two Intel(R) Xeon E5-2620 v2 CPUs (12-core, 2.6GHz), 16GB RAM and a GTX 780 graphics card. When dealing with recognition, keypoint detection and local feature extraction on the probe are conducted online only once, while the time cost in matching is multiplied by the number of the gallery samples (1-to-N matching), leading to the computation cost of 524ms (72ms+52ms+100×4ms) of a 1-to-100 system.

### 7.8 Summary and Discussion

According to the experimental results, the proposed method outperforms its counterparts, thus proving more discriminative to distinguish NIR dorsal hand vein images, which is supported by two principal theoretical foundations. On the one hand, it depends on recent investigations in optical health science [Bashkatov et al. 2011], presenting that people identification using hand vein images should not only focus on the vein network but also make use of these optical properties of the surrounding skin regions whose spectral response conveys subject-dependent inhomogeneous
composition of blood and various chromophores and pigments. On the other hand, based on the progress achieved in psycho-visual studies [Morgan 2011], we interpret dorsal hand vein images as surfaces whose geometric characteristics, i.e., plateau, cliffs, ridges, valleys, etc., capture the subject-dependent variations of absorption and scattering attributes and could be perfectly characterized through curvature related quantities in differential geometry.

From the experimental viewpoint, local feature-based methods, e.g., [Wang et al. 2010a, Tang et al. 2012, Huang et al. 2012a, Wang et al. 2012], outperform holistic techniques, such as PCA and LDA based subspace analysis [Chen et al. 2011], by a gap reaching more than 15 points on the NCUT Part A database. When we only focus on local feature-based methods, because we interpret hand vein images as surfaces to characterize their optical properties of hand skin as well as subcutaneous tissue in addition to vein network, the proposed approach employs multi-order (1st, 2nd, 3rd) differential quantities closely related to differential geometry and hence provides a more comprehensive description of geometric properties in comparison with several existing local based methods that only exploit single type of features. The performance is hence better than that of SIFT [Wang et al. 2012], LCP (an LBP variant) [Wang et al. 2010a], or even a hybrid one by combining local and global features [Zhu & Huang 2012].

Figure 7.17: Matching the keypoints across (a) scale variations and (b) translations (in both sub-figures, left column from top to bottom: Harris-Laplace, DoG on OGM₁, DoG on OGM₃, DoG on OGM₅, and DoG on OGM₇; right column from top to bottom: Hessian-Laplace, DoG on OGM₂, DoG on OGM₄, DoG on OGM₆, and DoG on OGM₈).
Additionally, the proposed approach employs the SIFT-like matching framework, and inherits the reputed robustness to in-plane rotation, scale changes, and translations, hence showing the potential to be competent in more difficult and complicated scenarios. Unfortunately, to the best of our knowledge, there is no publicly available database which contains these challenges. As such, we illustrate the robustness using artificial examples. For instance, Fig. 7.17 depicts the matching results across scale variation and translation. The two dorsal hand vein images are from the same subject in NCUT Part A. In (a), the left is of the original size while the right is resized to 90%. In (b), the left is fixed while half of the right is occluded due to translation. From the figure, we can see that even if in different scales or with large translations, the points detected on two images can still be correctly associated.
Chapter 8

Conclusions and Future Work

Contents

8.1 Contributions ......................................................... 163
8.1.1 3D face recognition based on principal curvature measures . . . 163
8.1.2 3D face recognition with heterogeneous cross-resolution data 164
8.1.3 Shape analysis based anti-spoofing face recognition . . . . . . 165
8.1.4 Hand-dorsa vein recognition .................................... 165
8.2 Perspectives of Future Work ..................................... 165
8.2.1 Remeshing method in heterogeneous cross-resolution face recognition .................................................. 165
8.2.2 Deep learning applied in 3D face recognition ................. 166

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 background, motivations, objectives, methodologies, and the contributions of this thesis. In Chapter 2, we extensively introduce the representative 3D face recognition approaches in three categories: the holistic feature based approaches, the local region-wise feature based approaches, and the local point-wise feature based approaches. The main research works are presented in the following three chapters. In Chapter 3, we introduce the principal curvature measures. In Chapter 4, we present the LPCMP feature descriptor based approach and the related 3D FR experiments. In Chapter 5, we introduce the PCM-meshSIFT feature descriptor and evaluate its performance in extensive experiments. In Chapter 6, we explore the anti-spoofing performance of the PCM-meshSIFT feature descriptor, and introduce the liveness detection based hand-dorsa vein recognition. In this last chapter, we conclude this thesis and list the perspectives of the future works.

8.1 Contributions

8.1.1 3D face recognition based on principal curvature measures

In general, a more stable and reliable recognition performance is the common objective in the research of 3D face recognition. For this objective, we propose to adopt
principal curvature measures in 3D face recognition. Through the study of principal
curvature measures, estimating these geometric quantities directly on the triangle
meshed model is their outstanding merit in 3D face recognition. It means that they
can describe correctly the shape of 3D meshed face scans. Motivated by the out-
standing property of principal curvature measures, we propose two approaches for
3D face recognition by designing two discriminative facial feature descriptor based
on these geometric measurements. The first feature descriptor is to encode the cur-
vature faces based on three principal curvature measures and enables accurate and
fast characterization of local shape. This approach achieves 93.16% rank-one recog-
nition rate on the whole FRGCv2 database. In order to handle the pose changes
and occlusion variations in 3D face recognition, the second registration-free feature
descriptor is proposed to construct a more effective and more robust system. This
system consists of the keypoints detection based on DoC, the assignment of the
canonical direction for the keypoint, and the keypoint description with three HOC
based descriptors. This approach raises the rank-one recognition to 98.17% on the
whole FRGCv2 database. Furthermore, the experiments involving the pose changes
and occlusion are also performed on the subsets of Bosphorus database for evaluat-
ing the robustness of proposed system. This approach achieves 92.75% recognition
rate in the pose changes subset and 98.43% recognition rate in the occlusion subset.

8.1.2 3D face recognition with heterogeneous cross-resolution data

For balancing the effectiveness against the efficiency of 3D face recognition system,
we propose to perform the heterogeneous cross-resolution based 3D face recognition.
The high-resolution (HR) face scans are enrolled into the gallery set, and the low-
resolution (LR) face scans count as the probe samples. The Signed Distance Func-
tion is utilized to merge frames of range images and to optimize the low-resolution
facial surface. In consideration of the promising performance of PCM-meshSIFT
feature descriptor in the homogeneous resolution based 3D face recognition, we con-
tinues to adopt it in this approach. Particularly, the keypoints are replaced by the
landmarks defined by Active Shape Models and the derivative interest points. This
modified system firstly performs the recognition experiments with the homogeneous
resolution data from Lock3DFace database. It obtain 96.24% recognition rate with
HR samples and 93.27% recognition rate of LR samples. In the heterogeneous cross-
resolution experiment, the 83.37% rank-one recognition rate demonstrates that the
feature descriptor is able to match the data in different resolutions with sufficient
reliability.

164
Chapter 8. Conclusions and Future Work

8.1.3 Shape analysis based anti-spoofing face recognition
To make full use of the shape characterization ability of the proposed feature descriptor, we exploit its performance of verifying the genuine face and distinguishing the manufactured mask in parallel. For this purpose, we design three kinds of the anti-spoofing evaluation related experimental scenarios: 1. Baseline evaluation of verification with Morpho database and FRGC v2.0 database, respectively; 2. Basic anti-spoofing performance evaluation with only Morpho Database; 3. The real-life simulating scenario by raising the ratio of the genuine face samples to the fraud mask samples with Morpho database and FRGCv2 database. The comprehensive results proves that our system guarantee competitive verification accuracy for genuine faces and promising anti-spoofing performance against spoofing mask attacks.

8.1.4 Hand-dorsa vein recognition
If the shape analysis based 3D face recognition method is an effective method to distinguish the masks from the genuine faces, the hand-dorsa vein recognition is the system to prevent all latent intrusive actions. The liveness detection based data acquisition pattern only works on the alive human body. In order to enhance the effectiveness of hand-dorsa vein recognition, we highlight the texture information existing not only in the vein vessels but also in the surrounding skin by using multiple order gradients. We locate the multi-source keypoints in the hand-dorsa vein images, generate the SIFT-based feature descriptor and perform the keypoint-wise matching to predict the identity. Evaluated by the comprehensive experiments, the proposed method achieves the best performance so far known on the NCUT Part A dataset, showing its considerable performance. Additional results on NCUT Part B also illustrate its generalization ability and robustness to low quality data.

8.2 Perspectives of Future Work
In this section, some possible extensive works and future research directions are presented as follows.

8.2.1 Remeshing method in heterogeneous cross-resolution face recognition
In the research of heterogeneous cross-resolution face recognition, according to the achieved results, there is some climbing potential of our proposed 3D face feature descriptor. But it is restricted by the disparity of the quality of the heterogeneous 3D meshed face scans in different resolutions. The difference of the vertices density enlarges the intra-class distance of feature. A possible solution is to remesh the high-
resolution face scans and the low-resolution face scans with the techniques presented in [Alyuz et al. 2010], [Li & Zhang 2007] and [Li et al. 2009]. The remeshing methods enable us to transform the heterogeneous cross-resolution matching problem to the homogeneous resolution matching problem. In our opinion, the recognition performance will be further improved, but the increasing time cost also needs to be considered.

8.2.2 Deep learning applied in 3D face recognition

Benefiting from deep learning methods [Bengio et al. 2013], especially the Convolutional neutral Network (ConvNet) [Krizhevsky et al. 2012][Wu et al. 2015], 2D face recognition has achieved significant advances in recent years. Since the groundbreaking work of DeepFace [Taigman et al. 2014], ConvNets have continuously set new records on the LFW benchmark [Huang et al. 2007] and even achieved the results beyond the human performance in [Schroff et al. 2015]. Until now, the deep learning has been studied in face representation [Sun et al. 2014], face alignment [Zhang et al. 2016a], face verification [Sun et al. 2013], face recognition [Parkhi et al. 2015], and heterogeneous face recognition [Narang & Bourlai 2016]. It is worthy to combine the deep learning technique and the shape analysis based feature in 3D face recognition. For example, the curvature faces carrying the geometric information proposed in this thesis are suitable to be analysed and further classified by ConvNets. Moreover, based on the deep learning method, it is possible to replace the image by the 3D meshed model as input, and perform the ConvNets based method on this meshed model. It leaves us a new direction of 3D face recognition in the future.
During my Ph.D studying, I have published three publications in international conferences, one in international workshop, and one in international journal. One journal paper is currently under review.

**International Conference:**


**International Workshop:**


**International Journal:**


**International Journal under Review:**

Bibliography


Bibliography


174
Bibliography


Bibliography


179


[Liu et al. 2002] Qingshan Liu, Rui Huang, Hanqing Lu and Songde Ma. **Face recognition using kernel-based fisher discriminant analysis.** In IEEE International Conference on Automatic Face and Gesture Recognition, pages 197–201, 2002. 9

181


Bibliography


[Tonchev et al. 2013] Krasimir Tonchev, Agata Manolova and Ihor Paliy. Comparative analysis of 3D face recognition algorithms using range image and
Bibliography


189


Bibliography

Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 775–779, 1997. 9


[Zhang et al. 2011] Zhiwei Zhang, Dong Yi, Zhen Lei and Stan Z Li. Face liveliness detection by learning multispectral reflectance distributions. In IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, pages 436–441, 2011. 121


