

# **Statistical Modelling of Craniofacial Shape**

Hang Dai

University of York

Computer Science

September 2018

*To my parents,  
Dai Hengliang and Zhang Xueping,  
and my wife,  
Liao Caiyun.*

# Abstract

With prior knowledge and experience, people can easily observe rich shape and texture variation for a certain type of objects, such as human faces, cats or chairs, in both 2D and 3D images. This ability helps us recognise the same person, distinguish different kinds of creatures and sketch unseen samples of the same object class. The process of capturing this prior knowledge is mathematically interpreted as statistical modelling. The outcome is a morphable model, a vector space representation of objects, that captures the variation of shape and texture. This thesis presents research aimed at constructing 3DMMs of craniofacial shape and texture using new algorithms and processing pipelines to offer enhanced modelling abilities over existing techniques. In particular, we present several fully automatic modelling approaches and apply them to a large dataset of 3D images of the human head, the *Headspace dataset*, thus generating the first public shape-and-texture 3D Morphable Model (3DMM) of the full human head. We call this the *Liverpool-York Head Model*, reflecting the data collection and statistical modelling respectively. We also explore the craniofacial symmetry and asymmetry in template morphing and statistical modelling. We propose a Symmetry-aware Coherent Point Drift (SA-CPD) algorithm, which mitigates the tangential sliding problem seen in competing morphing algorithms. Based on the symmetry-constrained correspondence output of SA-CPD, we present a symmetry-factored statistical modelling method for craniofacial shape. Also, we propose an iterative process of refinement for a 3DMM of the human ear that employs data augmentation. Then we merge the proposed 3DMMs of the ear with the full head model. As craniofacial clinicians like to look at head profiles, we propose a new pipeline to build a 2D morphable model of the craniofacial sagittal profile and augment it with profile models from frontal and top-down views. Our models and data are made publicly-available online for research purposes.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiii</b>
<b>Acknowledgements</b>	<b>xv</b>
<b>Declaration</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Modelling of Orthogonal Craniofacial Profiles . . . . .	6
1.2 3D Statistical Modelling Pipeline . . . . .	7
1.3 Symmetric Morphing and Symmetry-factored Statistical Modelling . . . . .	8
1.4 Modelling and Merging High Resolution Ear Shape . . . . .	11
1.5 Overview of Chapters . . . . .	13
<b>2 Literature Review</b>	<b>15</b>
2.1 Historical Perspective . . . . .	16
2.2 3D Face Models . . . . .	17
2.3 Ear Models . . . . .	20
2.4 3D Face/Head Database . . . . .	21
2.5 Facial Landmarking . . . . .	23
2.6 Correspondence Establishment . . . . .	26
2.7 Procrustes Alignment . . . . .	33
2.8 Statistical Modelling . . . . .	34

2.9	3D Facial Symmetry and Asymmetry . . . . .	36
2.10	Existing 3DMM Construction Pipelines . . . . .	38
2.11	Evaluation Criteria . . . . .	39
2.12	Summary . . . . .	41
<b>3</b>	<b>Modelling of Orthogonal Craniofacial Profiles</b>	<b>43</b>
3.1	Model Construction Pipeline . . . . .	44
3.2	Morphable Model Evaluation . . . . .	49
3.3	Summary . . . . .	53
<b>4</b>	<b>3D Statistical Modelling Pipeline</b>	<b>55</b>
4.1	Overview of Headspace Dataset . . . . .	56
4.2	Data Pre-processing . . . . .	57
4.3	Correspondence Establishment . . . . .	59
4.4	Similarity Alignment and Statistical Modelling . . . . .	68
4.5	Texture Map . . . . .	69
4.6	Evaluation of Correspondences . . . . .	72
4.7	Evaluation of 3DMMs . . . . .	77
4.8	Applications . . . . .	88
4.9	Summary . . . . .	94
<b>5</b>	<b>Symmetric Morphing and Symmetry-factored Statistical Modelling</b>	<b>97</b>
5.1	Notation and Preliminaries . . . . .	100
5.2	Symmetric Shape Morphing . . . . .	103
5.3	Evaluation of Symmetric Deformation . . . . .	111
5.4	Symmetry-factored Statistical Modelling of Craniofacial Shape . . . . .	116
5.5	Evaluation of Symmetry-factored Modelling . . . . .	120
5.6	Summary . . . . .	129
<b>6</b>	<b>Modelling and Merging High Resolution Ear Shape</b>	<b>131</b>
6.1	A 3DMM of the Ear . . . . .	132
6.2	Merging 3D Morphable Models . . . . .	137
6.3	Evaluation . . . . .	141

---

6.4	Summary . . . . .	149
<b>7</b>	<b>Conclusions</b>	<b>151</b>
7.1	Summary of Contributions . . . . .	151
7.2	Critical Analysis . . . . .	153
7.3	Future Work . . . . .	154
	<b>Appendix</b>	<b>157</b>
<b>A</b>	<b>List of Publications</b>	<b>157</b>
<b>B</b>	<b>Hierarchical Parts-based Template Morphing Framework</b>	<b>159</b>
B.1	Dense Correspondence . . . . .	161
B.2	Results . . . . .	163
	<b>References</b>	<b>165</b>



# List of Figures

1.1	A morphable model of triangles . . . . .	2
1.2	Template morphing for dense correspondence. . . . .	4
1.3	Liverpool-York Head Model. . . . .	5
2.1	Statistical modelling flow. . . . .	15
2.2	55 landmarks on ear . . . . .	21
2.3	2D facial landmarks detection systems. . . . .	24
2.4	Mixture-of-trees model . . . . .	26
3.1	The pipeline for 2D morphable model construction . . . . .	45
3.2	3D pose normalization using the texture information. . . . .	46
3.3	Compactness . . . . .	50
3.4	Specificity and generalization . . . . .	51
3.5	Separate models . . . . .	51
3.6	Three orthogonal profiles placed within a 3D frame. . . . .	52
3.7	Global profile model . . . . .	52
4.1	The flowchart of 3DMM construction pipeline. . . . .	55
4.2	Age distribution of subjects. . . . .	56
4.3	Pre-processing of 3D raw data . . . . .	58
4.4	Landmarks detection results . . . . .	60
4.5	Template morphing framework . . . . .	61
4.6	Proposed method compared with standard CPD . . . . .	62
4.7	Annotation of facial parts on template . . . . .	64
4.8	Template comparison with raw data . . . . .	64

4.9	Texture mapping flow . . . . .	70
4.10	Per-vertex color texture and texture map . . . . .	71
4.11	Internal comparison of approaches . . . . .	73
4.12	Improvement in correspondence and convergence . . . . .	73
4.13	Morphing results . . . . .	75
4.14	Self comparison . . . . .	76
4.15	Correspondence comparison . . . . .	76
4.16	Proportion of subjects with $<$ Euclidean distance error . . . . .	77
4.17	Global model of craniofacial shape . . . . .	78
4.18	Shape models . . . . .	79
4.19	Texture models . . . . .	80
4.20	t-SNE in terms of gender and age . . . . .	82
4.21	Compactness of sub-models. . . . .	82
4.22	Generalisation of sub-models. . . . .	83
4.23	Specificity of sub-models. . . . .	83
4.24	Compactness . . . . .	84
4.25	Generalisation and specificity . . . . .	84
4.26	Critical analysis of the proposed model . . . . .	85
4.27	The flexibility of reconstruction from incomplete data . . . . .	87
4.28	Age regression from 5 years to 80 years. . . . .	87
4.29	Registration results of patients . . . . .	89
4.30	BS intervention outcome . . . . .	90
4.31	TCR intervention outcome . . . . .	90
4.32	Case study . . . . .	91
4.33	Craniofacial anthropometric landmarks . . . . .	92
4.34	Craniofacial anthropometric measurements . . . . .	93
4.35	Craniofacial anthropometric measurements . . . . .	94
5.1	Symmetry-factored statistical modelling framework. . . . .	98
5.2	Why do we need symmetric deformation . . . . .	99
5.3	End results of each step . . . . .	100
5.4	Symmetry contour comparisons after template deformation . . . . .	104

---

5.5	Symmetry aware CPD process . . . . .	106
5.6	Deformation of the template . . . . .	112
5.7	3D morphable model constructed by SA-CPD . . . . .	113
5.8	Proportion of subjects with an error . . . . .	113
5.9	NPE and SCE against Gaussian noise . . . . .	114
5.10	Deformation results against outlier . . . . .	115
5.11	Visualisation of symmetry models . . . . .	121
5.12	Visualisation of asymmetry models . . . . .	122
5.13	Results of symmetry plane alignment, symmetrisation and model fitting .	123
5.14	Per-vertex fitting error and plot of eigenvalues . . . . .	124
5.15	Compactness and generalisation error of the proposed model and PCA model . . . . .	124
5.16	Symmetrisation comparison . . . . .	126
5.17	Simulation of half head completion from profile image. . . . .	127
5.18	Cumulative error distributions of the per-vertex fitting error from half head	128
6.1	Deformation to overcome over-fitting . . . . .	132
6.2	The flowchart of modelling and merging high resolution ear shape. . . . .	133
6.3	Iterative model construction process . . . . .	134
6.4	3D morphable model of ear . . . . .	135
6.5	The merging flowchart . . . . .	138
6.6	Joint template. . . . .	139
6.7	Results of each step in augmentation . . . . .	142
6.8	Augmentation results . . . . .	143
6.9	Average landmarks distance error for four system variants . . . . .	143
6.10	Ear model evaluation . . . . .	144
6.11	High resolution 3DMM of ear merged into a mean head . . . . .	145
6.12	High resolution 3DMM of ear merged into a head model . . . . .	146
6.13	Comparison of ear merging . . . . .	146
6.14	Ear merging results . . . . .	147
6.15	Fitting results of a single 3DMM and the proposed part-based 3DMM. . .	148

## List of Figures

---

B.1	Parts-based template morphing framework . . . . .	159
B.2	Deformation results . . . . .	163
B.3	Proportion of subjects with $<$ Euclidean distance error . . . . .	164

# List of Tables

2.1	3D face models . . . . .	20
2.2	List of 3D face/head databases. Sub.: subjects, Occl.: occlusions . . . . .	22
2.3	Registration methods . . . . .	27
2.4	Pipelines comparison . . . . .	39
3.1	Gender classification. . . . .	53
3.2	Age classification. . . . .	53
4.1	Comparison of 3DMM construction pipelines . . . . .	78
4.2	Texture map image quality assessment . . . . .	88
5.1	Gender classification results . . . . .	114
5.2	Age classification results . . . . .	116
5.3	Age classification using shape parameters . . . . .	127



# Acknowledgements

Foremost, I would like to begin by thanking my parents not only for their support throughout my educational career, but also for encouraging me to pursue my PhD. This is even tougher for my father who was diagnosed with an early-stage bladder cancer at the beginning of my PhD. I would like to dedicate this thesis to him, a tough fighter against cancer and super hero in my life.

Besides my parents, I would like to express my sincere gratitude to my supervisors and my friends Dr. Nick Pears and Dr. William Smith for the continuous support of my PhD study and research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. When I plan to go around UK, Nick always shows me the best part of the area. He provides me the chance to enjoy the best environment for academic research and everyday life in York. I would never forget the black sheep and blood pudding, the best combo recommended by Nick. Word can't describe how thankful I am. Will was my internal assessor for the first year of my PhD. Then he became my co-supervisor after a few interesting discussions. Every time I go to his office and ask 'May I grab a minute?', the answer is always 'Yes' with a gentle smile. I could not have imagined having better supervisors and friends. I would like to thank Professor Richard Wilson who, as my assessor, has provided a different perspective on my work. He has been through all the progress I've made in my PhD. His warm smile makes me relaxed and my brain function properly in the assessment meeting. I would like to thank Dr. Stefanos Zafeiriou for being my external examiner.

My sincere thanks also goes to my fellow labmates in research group: Dr. Chao Zhang, Dr. Chengliang Dai, Dr. Hao Sun, Savan Vachhani, Chaitanya Kaul and Dr. Antonio Arnone, for the stimulating discussions, and for all the fun we have had in the last three years. Also I thank my friends in Zhejiang University: Dr. Shaoze Wang, Hang

---

Zhang, Nan Li, Yifan Zhu, Hongda Chen and Xinyu Zhao. In particular, I am grateful to Professor Yong Ding for enlightening me the first glance of research.

Finally, I would like to thank my wife who always gives me unstinting support for my research and my life. We have been married for 4 years, which is not easy for a young couple when her visa application failed at the beginning of my PhD. Almost two years without her is a nightmare for me. Luckily, she joined me at the last year of my PhD. We've had so many brilliant memories in York that it is difficult for us to say goodbye to this amazing place. I feel lucky that I was offered a good job in academic institution when writing this thesis. I will further my research in 3D visualisation of tumour which aims at the early diagnosis of cancer.

# Declaration

I declare that the research described in this thesis is a presentation of original work, which I undertook at the University of York during 2015 - 2018. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Except where stated, all of the work contained within this thesis represents the original contribution of the author.

Some parts of this thesis have been published in conference proceedings and journals; where items were published jointly with collaborators, the author of this thesis is responsible for the material presented here.

Chapter 3 includes:

- H Dai, N Pears, C Duncan, 2D Morphable Model of Craniofacial Profile and Its Application to Craniosynostosis. Annual Conference on Medical Image Understanding and Analysis (MIUA), 731-742 2017.
- H Dai, N Pears, C Duncan, Modelling of Orthogonal Craniofacial Profiles. Journal of Imaging 3 (4), 55, 2017.

Chapter 4 includes

- H Dai, N Pears, WA Smith, Non-rigid 3D Shape Registration using an Adaptive Template. Proceedings of the European Conference on Computer Vision Workshop (ECCV-W) Munich, Germany, 2018.
- H Dai, N Pears, WA Smith, C Duncan, A 3D Morphable Model of Craniofacial Shape and Texture Variation. Proceedings of the International Conference on Computer Vision (ICCV) 2017.

- 
- B Robertson, H Dai, N Pears, C Duncan, A morphable model of the human head validating the outcomes of an age-dependent scaphocephaly correction. *International Journal of Oral and Maxillofacial Surgery*, vol 46, p68, 2017.

Chapter 5 includes:

- H Dai, N Pears, WA Smith, C Duncan, Symmetric Shape Morphing for 3D Face and Head Modelling. *IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*.
- H Dai, WA Smith, N Pears, C Duncan, Symmetry-factored Statistical Modelling of Craniofacial Shape. *Proceedings of the International Conference on Computer Vision Workshop (ICCV-W) PeopleCap 2017*.

Chapter 6 includes:

- H Dai, N Pears, WA Smith, A Data-augmented 3D Morphable Model of the Ear. *IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*.

**Copyright © 2018 by Hang Dai**

The copyright of this thesis rests with the author. Any quotations from it should be acknowledged appropriately.

# Chapter 1

## Introduction

*“The human face – in repose and in movement, at the moment of death as in life, in silence and in speech, when alone and with others, when seen or sensed from within, in actuality or as represented in art or recorded by the camera – is a commanding, complicated, and at times confusing source of information.”*

– Paul Ekman [58]

The theme throughout this thesis is how to model the 3D shape and texture variation of some object class. With prior knowledge and experience, people can easily observe rich shape and texture variation for a certain type of object, such as human faces, cats or chairs, in both 2D and 3D images. This ability helps us recognise the same person, distinguish different kinds of creatures and sketch unseen samples of the same object class. The process of capturing this prior knowledge is mathematically interpreted as statistical modelling. One such outcome is a morphable model, a vector space representation of objects, that captures the variation of shape and texture. Any convex combination of vectors of a set of object class examples generates a real and valid example in this vector space. Morphable models have many applications in creative media, medical image analysis and biometrics, by providing a useful encoding and prior statistical distribution of both shape and texture.

Figure 1.1 demonstrates a toy example of a morphable model using triangles. The training dataset is shown in red, the blue triangle is the mean shape, and the green triangle is an unseen example for the dataset. When representing a set of triangles in a vector

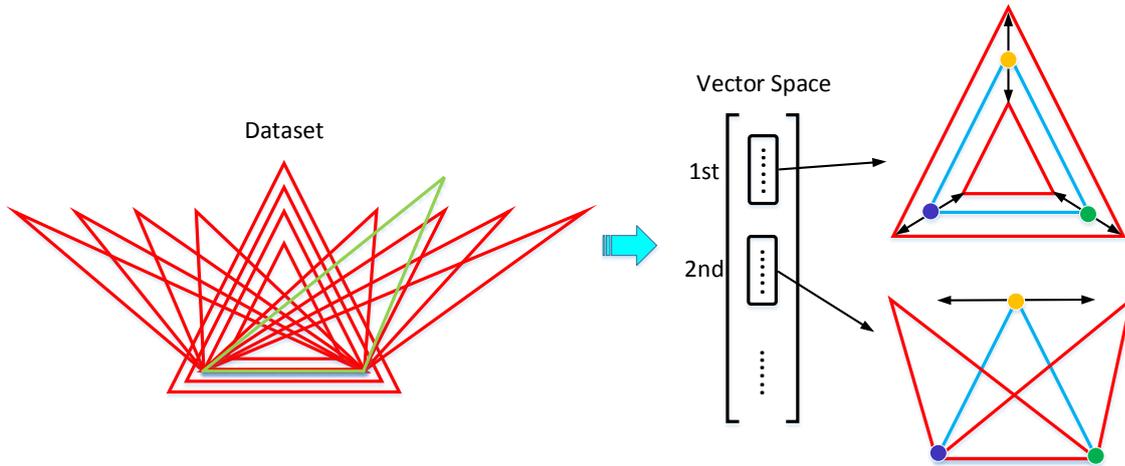


Figure 1.1: A morphable model of triangles. Red are the dataset shapes for training, blue is the mean training shape and green is an unseen example.

space, those vectors describe different types of shape variation: in our toy example, the first vector captures the size changing and the second vector captures horizontal motion of the yellow point. (In real examples, the variations may not be so easily described.) A certain linear combination of the first and second vectors of shape variation can represent a green triangle, an unseen example, with both size variation and horizontal motion of the yellow point. This new example is still within the same object class.

Statistical shape modelling is extensively studied in a variety of disciplines; for example computer vision, where researchers focus on applications in medical imaging, biometrics and the creative industries. In clinical usage, statistical shape analysis can help the surgeon to do planning, assessment, and follow-up of operations. Statistical shape modelling aims to characterise the mean shape, and the variances and covariances of different object parts for various classes of object.

Shape, as defined by D.G. Kendall [86], is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. In other words, those similarity effects need to be filtered out by aligning a collection of shape when doing shape analysis. A shape is described by locating a number of points on the outline. These points are defined as points of correspondence on each object that matches between and within populations. Statistical shape modelling is perhaps most commonly performed by Principal Component Analysis (PCA) over a set of meshes, which finds the directions in

---

the vector space that have maximum variance, whilst being mutually orthogonal. This process is feasible if and only if each mesh is reparametrised into a consistent form where the number of vertices, the triangulation, and the (approximate) anatomical meaning of each vertex are made consistent across all meshes. For example, given a vertex with index  $i$  in one mesh corresponding to the left mouth corner, it is required that the vertex with the same index in every mesh should correspond to the left mouth corner too. Meshes, every vertex of which satisfies the above properties, are said to be in dense correspondence with one another. This correspondence problem is easy to state informally, while it is challenging to solve accurately and robustly between highly variable meshes. The smooth area, such as cranium or cheek in our 3D head dataset, makes the objective measurement of correspondence quality more difficult than that in face only dataset.

There are various ways in which such a correspondence can be established. One promising approach is the template morphing method. Figure 1.2 describes the key in this approach, which is to morph the template to the example shapes in the training set. Template morphing brings shapes into dense correspondence by some combination of global and local transformation. Template-based correspondence methods need very high quality shape registration for good performance [151, 152]. When a template is moved and deformed  $N$  times to fit each of  $N$  objects in some class of 3D objects, the set of deformed templates has the same number of vertices and connectivity relationship as the (original) template.

Unfortunately high quality 3DMMs are comparatively hard to obtain and use. To train a model, a large and diverse set of high quality 3D scans are required. These scans need to be brought into dense correspondence with a template morphing algorithm. High quality data availability and non-trivial training are the main reasons for the limited adoption of 3DMMs, and there are very few readily-available 3DMM construction pipelines.

This thesis is concerned with 3D statistical shape modelling with particular reference to applications associated with craniofacial data. Statistical shape modelling is an essential component in many 3D shape analysis tasks. In this research, we attempt to build a mathematical and computational model of the human head - both face and cranium. Traditional model building pipelines have used manual landmarking to initialise surface alignment. However, this is extremely time consuming and laborious for large-

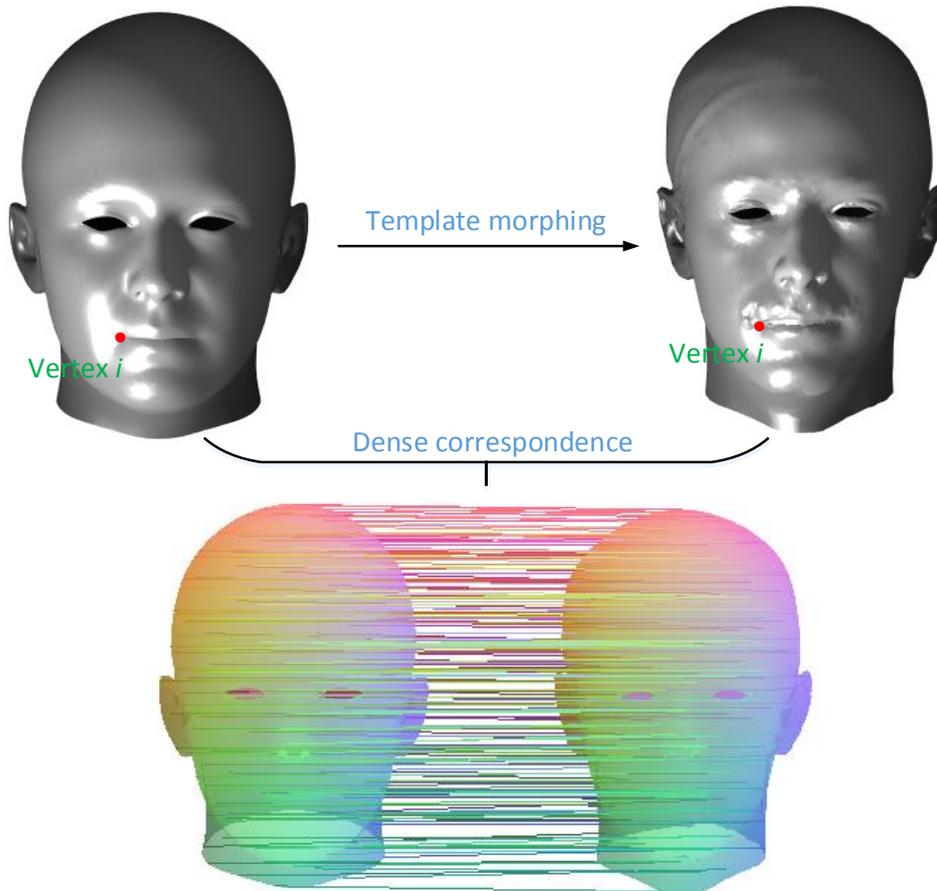


Figure 1.2: Template morphing for dense correspondence.

scale datasets. Here we present a fully automatic approach and apply it to a large dataset of 3D images of the human head, the *Headspace* dataset, thus generating the first public shape-and-texture 3D morphable model of the full human head: the Liverpool-York Head Model (LYHM). The model is named as such to reflect the data collection and model construction respectively. As illustrated in Figure 1.3, the Basel Face Model (BFM) [105] and Large Scale Face Model (LSFM) [27] are face models and the LYHM is the only full head model.

A full head model opens up new opportunities not afforded by existing 3D face models. For example: (i) we can reconstruct the most likely full head shape from partial data. This allows the modelling of different hair styles onto a correctly proportioned cranium (cranial shape usually cannot be directly observed), (ii) any part of the head shape can be matched against population norms thus facilitating tools for craniofacial diagnoses and surgical intervention evaluation, and (iii) the ability to complete partial views of the head

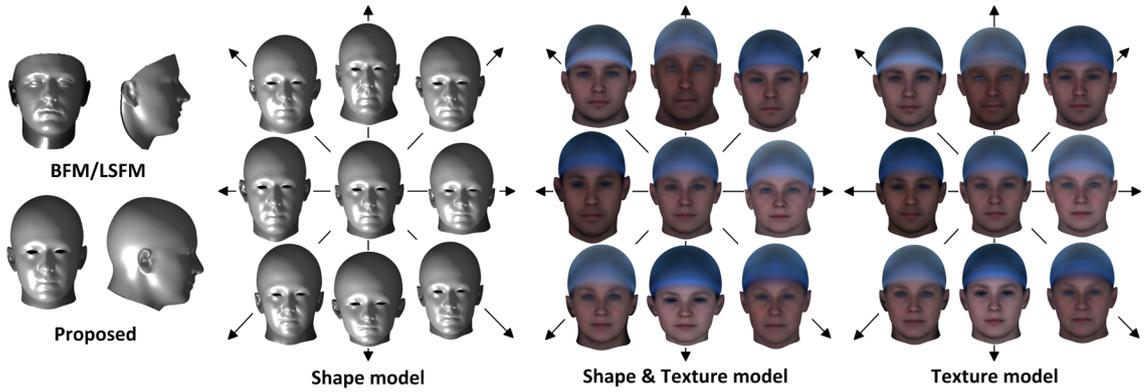


Figure 1.3: The proposed Liverpool-York Head Model. 1st block - shape compared to the Basel Face Model (BFM) [105] and Large Scale Face Model (LSFM) [27]; 2nd-4th blocks: the central head is the mean and the first four principal modes of variation are shown.

may be exploited in biometric applications.

In the following four subsections, we introduce the four main themes presented in the main body of the thesis, but first we motivate these. At the beginning of the research, we treated the 2D morphable model construction as a training exercise for the full 3D modelling. This familiarised us with the whole pipeline of morphable model construction. Furthermore 2D modelling is useful in itself. It is often useful, in terms of visual clarity and attention focus, for the craniofacial clinician to examine shape outlines from canonical viewpoints. After gaining familiarity with the morphable model construction pipeline, we moved onto 3DMM construction. We proposed a fully automatic 3DMM construction pipeline with contributions in pose normalisation method and correspondence establishment method. We used this pipeline to build the first 3DMM of the full head. When building this 3DMM, we noticed both the symmetry contour sliding problem in the existing correspondence establishment methods, and the asymmetric shape variation in the model. In order to overcome the symmetry contour sliding problem and analyse the asymmetric shape variation, we exploited the craniofacial symmetry and asymmetry in template morphing and statistical modelling. Moreover, we noticed that the ear mesh quality was not high enough. Therefore, a separate 3DMM of the ear was necessary to represent the actual ear shape. So we used high quality 3D ear meshes from CT scans with data augmentation to build a 3DMM of the ear. We then merged this ear model with

the proposed head model.

## 1.1 Modelling of Orthogonal Craniofacial Profiles

In the medical analysis of craniofacial shape, the visualisation of 2D profiles is highly informative when looking for deviations from population norms. It is often useful, in terms of visual clarity and attention focus, for the clinician to examine shape outlines from canonical viewpoints; for example, pre- and post-operative canonical profiles can be overlaid. We view profile-based modelling and analyses as being complementary to that of a full 3D shape model. Thus we present a pipeline that models 2D craniofacial profiles in Chapter 6.

Profile visualisations should be backed up by quantitative analysis, such as the distance (in standard deviations) of a patient’s shape profile from the mean profile of a reference population. Therefore, we have developed a novel image processing pipeline to generate a 2D morphable model of craniofacial profiles from a set of 3D head surface images. We construct morphable 2D profile models over three orthogonal planes to provide comprehensive models and analyses of shape outline.

The morphable model of orthogonal craniofacial profiles provides two functions. Firstly, it is a powerful prior on 2D profile shapes that can be leveraged in fitting algorithms to reconstruct accurate and complete 2D representations of profiles. Secondly, the proposed model provides a mechanism to encode any 2D profile in a low dimensional feature space; a compact representation that makes tractable many 2D profile analysis problems in the medical domain.

In summary, the contributions of this work is as follows: We propose a new pipeline to build a 2D morphable model of the craniofacial sagittal profile and augment it with profile models from frontal and top down views. We also integrate all three profiles into a single model, thus capturing any correlations within and between the three profile shapes more explicitly and clearly than is possible with PCA analysis on a full 3D model. Extensive qualitative and quantitative evaluations reveal that the proposed normalisation achieves state-of-the-art results.

## 1.2 3D Statistical Modelling Pipeline

Non-rigid 3D shape registration is the key component in our 3D statistical modelling pipeline. This is also known as template morphing, and is a key component in many tasks, such as 2D/3D image registration, morphable model construction and shape recognition. The goal of non-rigid shape registration is to align and deform (morph) a *source* point set to a *target* point set. By using some form of template shape as the source, morphing is able to reparametrise a collection of raw 3D scans of some object class into a consistent form. This facilitates full dataset alignment and subsequent 3DMM construction. In turn, the 3DMM constitutes a useful shape prior in many computer vision tasks, such as recognition and missing parts reconstruction.

Currently, methods that deform a 3D template to all members of a specific 3D object class in a dataset use the same template shape. However, datasets representative of global object classes often have a wide variation in terms of the spatial distribution of their constituent parts. Our object class in this thesis is that of the human face/head, where the relative positions of key parts, such as the ears, mouth, and nose are highly varied, particularly when trying to build 3DMMs across a wide demographic range of age, gender and ethnicity. Using a single template shape means that often key parts of the template are not at the same relative positions as those of the raw 3D scan. This causes slow convergence of shape morphing and, worse still, leads to end results that have visible residual errors and inaccurate correspondences in salient local parts.

To counter this, we propose an adaptive template approach that provides an automatically tailored template for each raw 3D scan in the dataset. The adaptive template is obtained from the original template using *sparse* shape information (typically point landmarks), thereby locally matching the raw 3D scan very specifically. Although this is a pre-process that involves template shape adaptation, we do not consider it as part of the main template morphing process, which operates over *dense* shape information.

In summary, the contributions of this work is as follows:

- We propose a fully automatic pipeline to build 3DMMs, with contributions in pose normalisation and dense correspondence - in particular we proposed a fully automatic registration framework using an adaptive template. We compared the pro-

posed methods with our previous methods and other competing algorithms;

- We present a high quality texture map from several views of cameras for the morphed template and use the texture map for texture modelling. This technique improves the texture image quality, which is validated by traditional image quality assessment methods.
- We build both global craniofacial 3DMMs and demographic sub-population 3DMMs from more than 1200 distinct identities in the *Headspace* dataset [1, 57] and we make both 3DMMs and Headspace dataset publicly available <sup>1</sup>. To our best knowledge, our models are the first public shape-and-texture craniofacial 3DMMs of the full human head;
- We demonstrate flexibility modes of our model such that, when given a fixed face shape, we compute the range of possible cranial shapes and vice-versa. We use a linear regression between the shape and texture parameters against age to observe the influence of age on craniofacial growth.
- We demonstrate the first clinical use of craniofacial 3DMMs in the assessment of two different types of surgical intervention applied to the crania. We also provide a fully automatic way to do facial anthropometric measurements.

### 1.3 Symmetric Morphing and Symmetry-factored Statistical Modelling

In chapter 4, we present techniques to analyse craniofacial symmetry and asymmetry. Most biological objects, including human heads and bodies, possess approximate symmetries. Often this is principally extrinsic, bilateral symmetry (i.e. reflective symmetry about a vertical plane bisecting the object). Deviations from exact symmetry are an interesting and potentially important geometric property in terms of modelling and understanding 3D shape variation. The degree of asymmetry may convey information about an object.

---

<sup>1</sup><https://www-users.cs.york.ac.uk/~nep/research/>

For example, there is evidence that facial asymmetry is used by humans to measure genetic health [81] and asymmetry in man-made objects may indicate imperfections in the manufacturing process. For this reason, we suggest that statistical shape models should separate symmetric from asymmetric shape variability in order to reveal interesting and potentially subtle aspects of shape variation over population datasets.

Our symmetric deformation algorithm builds on Myronenko and Song [100], who derived the Coherent Point Drift (CPD) point registration algorithm in the context of both global affine deformations (CPD-affine) and local non-rigid deformations (CPD-nonrigid). They noted that non-rigid point set registration is an ill-posed problem and to obtain a unique solution, constraints on the solution space are required. They regularised the non-rigid motion field using a Gaussian kernel, so that deformation varies smoothly ('drifts' coherently) over the template surface.

Our hypothesis is that a restriction to symmetric deformations may improve template morphing processes for (near) symmetric shapes; for example, it will not be possible for the sagittal symmetry contour of the template to deform via shearing and tangential surface sliding, which can occur in CPD-affine and CPD-nonrigid respectively. If the shape that we are modelling has an *approximate* reflective symmetry, we can use this as a powerful constraint that can be leveraged in *any* morphable model construction pipeline. Thus our contribution is a new 3D shape template morphing algorithm that is suitable for any class of shapes that has regularity associated with a plane of reflective symmetry. The human face and full head are examples of such classes. Our approach requires strong but not perfect symmetry, as a final stage in our algorithm morphs the symmetrically deformed template to any asymmetries in the data. Therefore, our algorithm also permits the decomposition of shape into symmetric and asymmetric components, which is an interesting aspect of the study of shape variations and covariations within datasets.

We call our method *Symmetry-aware CPD* (SA-CPD). In SA-CPD, we symmetrise both CPD-affine and CPD-nonrigid. In the global affine case, we remove any non-symmetric shearing operations. In the local non-rigid case, we find the nearest symmetric deformation, in a least squares sense, to that generated by CPD-nonrigid. These symmetric deformations allow us to generate a symmetric template shape that is usually very close to the data. Finally, to account for shape detail in the data, the template uses a

Laplace-Beltrami regularised projection operation.

The asymmetric variation in human head shape is much smaller than symmetric variation. The symmetric deformation algorithm facilitates the extraction of the symmetric and asymmetric shape variation. Classical models obtained by applying PCA directly to head shapes pools asymmetric variation with the much larger symmetrical variation. In this thesis, we consider how to build a symmetry-factored statistical shape model in a principled way. We address the problem with a symmetry-factored shape modelling pipeline and a global description of the resulting morphable model.

Once built, the symmetry-factored statistical modelling of craniofacial shape opens up new opportunities not afforded by existing 3D morphable models. Firstly, it provides a different way to reconstruct a full head model from a profile scan beyond just using the reflection of the profile. Secondly, we can analyze the asymmetry of multiple craniofacial regions. Thirdly, we can investigate the different importance of symmetry-factored information in different applications, e.g. face recognition or ergonomics.

In summary, the contributions of this work is as follows:

- We propose a shape template morphing approach suitable for any class of shapes that exhibits approximate reflective symmetry over some plane. The human face and full head are examples. A shape morphing algorithm that constrains all morphs to be symmetric is a form of deformation regulation. This mitigates undesirable effects seen in standard morphing algorithms that are not symmetry-aware, such as tangential sliding.
- Our method builds on the Coherent Point Drift (CPD) algorithm and is called Symmetry-aware CPD (SA-CPD). Global symmetric deformations are obtained by removal of asymmetric shear from CPD's global affine transformations. Symmetrised local deformations are then used to improve the symmetric template fit. These symmetric deformations are followed by Laplace-Beltrami regularized projection which allows the shape template to capture the shape detail in the raw shape data. The pipeline facilitates construction of statistical models that are readily factored into symmetrical and asymmetrical components. Evaluations demonstrate that SA-CPD mitigates tangential sliding problem in CPD and outperforms other

competing shape morphing methods, in some cases substantially.

- 3D morphable models are constructed from over 1200 full head scans, and we evaluate the constructed models in terms of age and gender classification. The best performance, in the context of SVM classification, is achieved using the proposed SA-CPD deformation algorithm;
- Based on the symmetry constrained correspondence output of SA-CPD, we present a symmetry-factored statistical modelling method for craniofacial shape. Our main contribution is to show how to build a statistical model with separate parameters for symmetric and asymmetric variations.
- The resulting model is still linear and so can be used in place of any existing 3DMM but with the additional ability to separate symmetric from asymmetric variation. This include a method for symmetrisation regularised by the Laplace-Beltrami operator, symmetry-aware Generalized Procrustes Analysis (GPA) and the symmetry-factored statistical modelling method.
- We make available, for the first time, a morphable model of craniofacial symmetry and asymmetry. Comprehensive evaluation shows that the proposed model has significantly better performance than the standard methods.

## 1.4 Modelling and Merging High Resolution Ear Shape

In chapter 5, we will present research on part-based 3D morphable models, with particular emphasis on modelling the complex shape of the ear as a separate part.

The shape of the ear has long been recognised as a means of biometric identification. Morphable models provide powerful statistical priors on shape and so can be used in biometric ear analysis. We present a pipeline capable of building a 3D morphable model of the human ear from a very limited training sample of 3D ears, using data augmentation.

Deep learning can be used on 3D meshes, but a large-scale dataset is needed for training and it takes a huge effort to capture a sufficiently large number of 3D meshes. Since 3D data augmentation can generate new 3D data, it satisfies the need for a large dataset and has become an active research field. Most 3D data augmentation methods employ

Convolutional Neural Networks (CNNs) to generate the 3D data as realistically as possible. However, these works rely on an adequate morphable model. Thus, CNN-based methods are not applicable with an inadequate morphable model i.e. one built from a limited amount of training data.

We have 20 high quality 3D meshes of the ear [160], taken from 10 subjects, with the left ear reflected to be compatible with the right ear shape. This is insufficient to construct a 3D morphable model that is a good representation of the mean ear shape and the variance and covariances of size-and-shape (form), over a large population. However, with such a limited dataset, we construct an initial approximate model of the ear. The model has over 7K vertices (7111) and we employ a modified version of our morphing technique (Dai et al. [47]) to build the model, which is an extension of Coherent Point Drift (CPD) [100]. Subsequently, 3D data augmentation is able to generate new samples for the 3DMM construction, thereby boosting the initial morphable model in terms of its accuracy in representing larger populations. Recently, Zhou et al. made a 2D ear image dataset available with 55 ground-truth landmarks [155] over 600 images, partitioned into 500 training images and 100 test images. We leverage the large annotated 2D ear dataset [155] to generate a large 3D ear dataset in the proposed data augmentation process. Our process includes the following stages: 1) landmark-based 3DMMs fitting; 2) use of 3D deformation to overcome the over-fitting (caused by an insufficient number of training subjects); 3) 3D mesh editing, regularized by 2D image information.

Ear detail cannot be captured easily by a 3D head scanner. Thus there is not much shape variation in the ear part of 3D head model. It is very useful to merge the ear model with the head model. To merge two morphable models, the expectation is that any sample meshes generated by one model can be perfectly merged into any sample meshes generated by the other one. The standard method is to use patch smoothing to move the noise data on the joint area between face and ear. However, this method omits the fact that the edges of the face-without-ear and the ear part always mismatch badly. So it often ends up with a discontinuous area, which is not a desirable outcome. To counter this, we firstly use a shape alignment to align the ear mesh to the ear region of face mesh. Then we use mesh manipulation to match the edges of the two separate parts. This is followed by a mesh smoothing method to remove most of the small residual noise. The results show

that the proposed method removes discontinuous areas in the merging process.

In summary, the contributions of this work is as follows:

- We present an iterative process of refinement for a 3D Morphable Model (3DMM) of the human ear that employs data augmentation. The process employs the following stages 1) landmark-based 3DMM fitting; 2) 3D template deformation to overcome noisy over-fitting; 3) 3D mesh editing, to improve the fit to manual 2D landmarks.
- These processes are wrapped in an iterative procedure that is able to bootstrap a weak, approximate model into a significantly better model. Evaluations using several performance metrics verify the improvement of our model using the proposed algorithm;
- We use this new 3DMM model-booting algorithm to generate a refined 3D morphable model of the human ear, and we make this new model and our augmented training dataset public;
- We merge the proposed 3DMMs of ear with the full head model, which ends up with a part-based morphable model of the full head. This part-based morphable model provides more shape variation and shape detail of the ear.

## 1.5 Overview of Chapters

The remainder of this thesis is arranged into the following chapters. In Chapter 2 we thoroughly review the relevant literature. This is necessarily broad ranging, since this thesis is concerned with topics from a number of fields: 3D face datasets, correspondence establishment, Procrustes alignment and statistical modelling. The review focuses on 3D morphable models of the face. We also pay special attention to 3D facial symmetry and asymmetry, which are explored in both correspondence establishment and statistical modelling.

Chapters 3-6 are the main body of the thesis, covering (i) orthogonal profile modelling, (ii) our 3D statistical modelling pipeline, (ii) symmetrical morphing and modelling, and (iv) modelling and merging ear shape, as described in Sections 1.1-1.4 above.

Finally, in Chapter 7, we review the contributions made in the thesis, highlight weaknesses in the work as it stands and suggest future areas for consideration and ways in which the work could be extended.

# Chapter 2

## Literature Review

In this chapter, we provide a thorough review of relevant literature, initially taking a broad historical perspective on shape analysis and modelling that pre-dates Blanz and Vetter’s seminal work on 3D Morphable Models (3DMMs) [20]. The review then focuses specifically on models of the facial region, giving examples of face and ear models in Section 2.2 and Section 2.3 respectively. Constructing a 3DMM is to set up a unified mathematical expression with a set of shape parameters. The general steps to build a 3DMM are shown in Figure 2.1. According to these steps, the review covers 3D face/head datasets (Section 2.4), correspondence establishment including both sparse data (i.e. landmarking for initialisation, Section 2.5) and dense data (Section 2.6), Procrustes alignment (Section 2.7) and statistical modelling (Section 2.8). In Section 2.9, we pay special attention to 3D symmetry and asymmetry, which are explored in both correspondence establishment and statistical modelling. In Section 2.10, we review existing 3DMM modelling pipelines, while the following section describes how to evaluate them. A final section is used to summarise the review.



Figure 2.1: Statistical modelling flow.

## 2.1 Historical Perspective

This section briefly highlights some key shape modelling developments that pre-date modern 3DMMs. Early work in shape analysis was performed on 2D images. In 1942, Thompson et al. [130] sketched transformation grids to show how one shape had to be deformed to match another in terms of their corresponding landmarks. Kendall [85] derived statistical developments concerning the shape space for Procrustes-registered data in 1984. Bookstein's work on Thin Plate Splines (TPS) [24] was central to the development of the related field of statistical shape analysis [56], which provides the theoretical underpinnings of the statistical shape modelling. Later Bookstein employed a combination of Procrustes analysis and TPS in order to analyse shapes in terms of their landmark positions [25]. Meanwhile, relevant work was produced in Computer Vision. For example, rigid transformation estimation between a pair of 3D shapes was solved by several Computer Vision researchers including approaches based on Singular Value Decomposition (SVD) that lead to least squares solutions are particularly popular [9].

In the 1990s, two groups of researchers Besl and McKay [18] and Chen and Medioni [36] independently proposed the Iterative Closest Points (ICP) algorithm for rigid shape alignment and registration. ICP cycles through three main steps: i) finding surface correspondences as closest points; ii) computing the rigid transformation estimation between them, and iii) applying the transformation to one of the surfaces to make it closer to the other. Later several research groups investigated non-rigid registration approaches where, for example, local affine deformations are permitted [5, 8].

Also in the 1990s, Cootes et al. developed shape models applied to 2D images, termed Point Distribution Models (PDMs) [41]. The work is done with reference to 2D shapes, where corresponding points are manually marked on the boundaries of a set of training examples. The points are aligned to minimise the variance in distance between corresponding points. This is done by encapsulating a Procrustes-based alignment in an iterative procedure, where the mean is normalised to a default scale and pose at each iteration. After such alignment, a standard principal component analysis (PCA) captures how the shapes deviate from the mean shape. Cootes et al. presented Active Shape Models (ASM) in [42], where pose, scale and shape parameters are determined in order to fit the model to an image. This work was inspired by the earlier work on active contour models [83].

The same research team also went on to include texture in their models to give active appearance models [40]. They developed a set of shape modelling approaches where the best correspondences are those that define the most compact shape model given some quality of fit between the model and the data [50, 91]. Terzopoulos and Metaxas [128] introduced a physically-based approach to fitting 3D shapes. They formulated deformable superquadrics which incorporate the global shape parameters of a conventional superellipsoid with the local degrees of freedom of a spline. Kakadiaris et al. [82] presented an integrated approach to do shape segmentation and motion estimation using a physics-based framework.

## 2.2 3D Face Models

Existing 3D statistical face models mainly consist of either morphable models, multilinear models and part-based models, as shown in Table 2.1. In the late 1990s, Blanz and Vetter built a 3DMM from 3D face scans [20] and employed it in 2D face recognition [21]. Two hundred scans were used to build the model (young adults, 100 males and 100 females). Dense correspondences were computed using optical flow with an energy term dependent on both shape and texture.

There are very few publicly-available morphable models of the human face and, to our knowledge, none that include the full cranium. The Basel Face Model (BFM) is the most well-known and widely-used and was developed by Paysan et al. [105]. Again 200 scans were used, but the method of determining corresponding points was improved. Instead of optical flow, a set of hand-labelled feature points is marked on each of the 200 training scans. The corresponding points on a template mesh are known, which is then morphed onto the training scan using under-constrained per-vertex affine transformations, which are constrained by regularisation across neighbouring points [8]. The technique is known as optimal-step Non-rigid Iterative Closest Points (NICP).

The Basel Face Model (BFM) [79] was released as both a global model and a part-based model that is learned for four regions (eyes, nose, mouth and the rest of face). In the part-based version, the regions are fitted to the data independently and merged in a post-processing step. The part-based model was shown to lead to a higher data accuracy

than the global model. This method and its follow-up work [15, 127] are implemented by:

- Manual face segmentation into several local regions;
- Morphable model construction for each segmented part;
- A post-processing step to stitch the segmented parts.

De Smet et al. [52] proposed a method to find the optimal segmentation automatically by clustering the vertices, which is based on features derived from their displacements. In order to address the potential discontinuities at the boundaries of the segments, they smoothly weight the segments to obtain regionalised basis functions for the training data.

A statistical model called the multi-linear model [22, 136, 146, 147] is employed to statistically model the varying facial expressions. By using a multi-linear model, Vlastic et al. [136] modelled facial shape using a combination of identity and expression variation. Yang et al. [147] modelled the expression of a face in a different input image of the same subject. A number of PCA shape spaces for each expression are built and combined with a multi-linear model. A follow-up work [22, 146] used this model for a better description of expressions in videos. When a sequence of 3D meshes is given, Bolkart et al. [22] fitted a multi-linear model to parametrise a 4D sequence. In 2015, they demonstrated a direct construction of multi-linear model from a set of meshes using a global optimization of 3DMM parameters along with a group-wise registration over the 3D scans [23]. Another alternative to modelling faces with expression is to blend different shape models with expressions, which was introduced by Salazar et al. [113] to establish correspondence among faces with expression.

A hierarchical pyramids method was introduced by Golovinskiy et al. to build a localised model [69]. In order to model the geometric details in a high resolution face mesh, this statistical model is able to describe the varying geometric facial detail. Brunton et al. [31] described 3D facial shape variation at multiple scales using wavelet basis. The wavelet basis provided a way to combine small signals in local facial regions which are difficult for PCA to capture.

Claes et al. [37] explored the independent effects of the sex, genomic ancestry and genotype on facial shape variation. The experimental results showed that a set of 20 genes has significant effects on facial shape variation.

In 2017, Booth et al. [27] built a Large Scale Facial Model (LSFM), using the NICP template morphing approach, as was used in the BFM, but with error pruning, followed by Generalised Procrustes Analysis (GPA) for alignment, and PCA for the model construction. This 3DMM employs the largest 3D face dataset to date, and is constructed from 9663 distinct facial identities.

Marcel et al. [94] model the shape variations with a Gaussian process, which they represent using the leading components of its Karhunen-Loeve expansion. This Gaussian Process Morphable Models (GPMMs) unify a variety of non-rigid deformation models with B-splines and PCA models as examples. In their follow-on work, they present a novel pipeline for morphable face model construction based on Gaussian processes [67]. GPMMs separate problem-specific requirements from the registration algorithm by incorporating domain-specific adaptations as a prior model.

Tran et al. [131] proposed a framework to construct a nonlinear 3DMM model from a large set of unconstrained face images, without collecting 3D face scans. Specifically, given a face image as input, a network encoder estimates the projection, shape and texture parameters. Two decoders served as the nonlinear 3DMM to map from the shape and texture parameters to the 3D shape and texture, respectively.

Tan et al. [125] employed mesh variational auto-encoders to explore the probabilistic latent space of 3D meshes. The training is performed on the mesh directly rather than the UV parameterization for the mesh. Genova et al. [66] presented a method for training a regression network from image pixels to 3D morphable model coordinates, where supervised training data is not necessary. Tewari et al. [129] fused a convolutional encoder with a differentiable renderer and a self-supervised training loss in a end-to-end training framework. Kim et al. [88] employed a deep convolutional inverse rendering framework for faces that aimed at estimating facial pose, shape, expression, reflectance and illumination, by estimating all parameters from just a single image.

### **2.2.1 Critical Analysis**

Over the last 20 years, research in 3DMM construction has mainly been focused on two main techniques: 1) dense correspondence establishment and 2) statistical modelling methods. The theme throughout this thesis is to develop better correspondence establish-

Table 2.1: 3D face models

Basis functions	Methods
Global	Morphable model (PCA) [7, 20]
Global	Multilinear model [22, 27, 136, 146]
Part-based	Part-based model [15, 52, 127]
Localised detail	Hierarchical pyramids [69]
Local	Local wavelet model [31]

ment algorithms and statistical modelling methods. Most of the methods need landmarks in either the initialization or the correction of correspondence establishment. Even some local models require manual division of the shape into several parts. So most methods are not fully automatic over the whole procedure.

## 2.3 Ear Models

The shape of the ear has long been recognised as a means of biometric identification [2, 108]. There are many existing ear recognition systems, with a recent survey by Emeršič et al. [60] and ear biometrics continues to be an active research area [59]. Morphable models provide powerful statistical priors on shape and so can be used in biometric ear analysis.

Zolfaghari et al [160] described the only construction of a morphable model for external ear shapes based on a deformation framework using diffeomorphic metric mapping. They release high quality 3D meshes of the ear for 10 subjects [160]. This is insufficient to construct a 3D morphable model that is a good representation of the mean ear shape and the variance and covariances of size-and-shape (form), over a large population.

Recently, Zhou et al. made a 2D ear image dataset available with 55 ground-truth landmarks [155] over 600 images, partitioned into 500 training images and 100 test images. Figure 2.2 shows the 55 landmarks and their semantic annotations. These landmarks may be used for data augmentation to reconstruct the outer ear without losing the topology information.

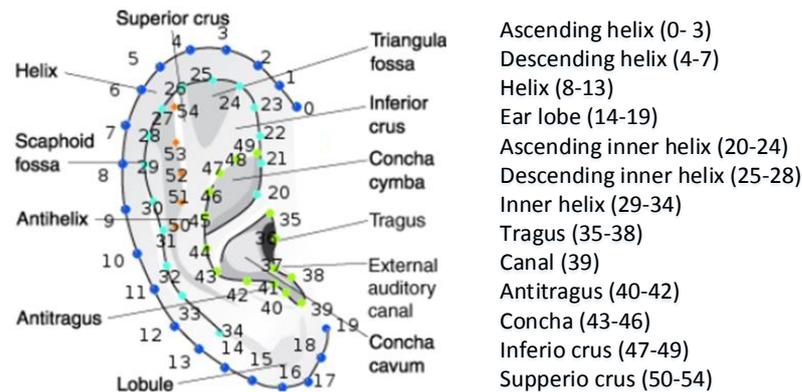


Figure 2.2: 55 landmarks on ear and their semantic annotations taken from [155].

### 2.3.1 Critical Analysis

High quality 3D ear data is very limited in its public availability. To our knowledge, Zolfaghari et al. [160] published the only work on a 3D morphable model of ear, but the morphable model is not publicly available. Zhou et al. [155] built a 2D morphable model of the ear and made a 2D ear image dataset publicly available with 55 manually-labelled ground-truth landmarks [155]. A large 3D ear dataset is desirable to build a representative 3D morphable model of the ear, but a high quality 3D ear is very difficult to collect directly. It is possible to leverage the large annotated 2D ear dataset [155] to generate a large 3D ear dataset in some form of data augmentation process.

## 2.4 3D Face/Head Database

The form of a 3D face dataset is application dependent. Most face databases are face recognition oriented. Only the Bosphorus dataset contains facial occlusions. FRGC v.2 [109] used to be the largest 3D face database. However, in 2016, MeIn3D replaced FRGC v.2 as the largest. The BU-3DFE [150] database has various emotional expressions. Every subject displays four intensity levels in each of six emotions. Table 2.2 lists the 3D face databases used in research. Not all of them are publicly available. Note that our database is new and is called the *Headspace* dataset and this is the first publicly-available full head database.

The Face Recognition Grand Challenge (FRGC) has been organized for several rea-

Table 2.2: List of 3D face/head databases. Sub.: subjects, Occl.: occlusions

Database	Sub.	Total	Expression	Pose	Occl.
FRGC v.2 [109]	466	4007	Anger, happiness, sadness, surprise, disgust, puffy	NA	NA
BU-3DFE [150]	100	2500	Anger, happiness, sadness, surprise, disgust, fear	NA	NA
ND2006 [62]	888	13450	Happiness, sadness, surprise, disgust, other	NA	NA
York [76]	350	5250	Happiness, anger, eyes closed, eye-brows raised	NA	NA
CASIA [154]	123	1845	Smile, laugh, anger, surprise, closed eyes	NA	NA
GavabDB [99]	61	549	Smile, frontal accentuated laugh, frontal random gesture	Left, right, up, down	NA
3DRMA [19]	120	720	NA	Slight left/right and up/down	NA
Bosphorus [114]	105	4652	34 expressions (action units and six emotions)	13 yaw, pitch and cross rotations	4 occl.
MeIn3D [27]	9663	12000	NA	NA	NA
FaceBase [117]	3500	3500	NA	NA	NA

sons. There were three main directions to be pursued : 1) high resolution still images, 2) multiple still images both on the enrollment and challenge side, 3) 3D face shape. Such databases were a challenge for teams in industry and academia. Since this dataset has various expression, it is very challenging to solve landmark localisation in FRGC. However, the mesh quality and resolution is relatively lower than those captured by the latest techniques.

The MeIn3D [27] database includes over 10,000 3D facial scans. It was collected by ibug group in Imperial London College in the Science Museum. Booth et al. [27] used

this dataset to construct a large scale 3D morphable model of face. They aimed at 3D face reconstruction and surgical planning. However, this dataset is face-only and not made public yet. The 3D mesh in this dataset composes of about 60,000 vertices joined into 120,000 triangulated faces associated with a high resolution texture image. Meta-data, including gender, age and ethnicity, is provided to build specific demographic model and analyze the shape variation for certain demographic group.

The FaceBase [117] dataset collects 3D face meshes along with DNA information. This dataset includes 35 age groups from 3 to 40 years old. It aims at collecting 50 males and 50 females in each age group. They released their measurements of craniofacial shape. This dataset has the potential to explore the relation between craniofacial shape and gene information.

### **2.4.1 Critical Analysis**

The MeIn3D and FaceBase dataset for craniofacial researchers contain a large amount of 3D face scans, but the full cranium is not included. The only other full head (including full cranium) dataset that we are aware of is that of the Size China project, but there is no 3DMM developed from Size China data. It is expensive for researchers to get access to Size China. Meanwhile, our Headspace dataset is the only full head dataset publicly available and free for researchers, which is why this work is the first to generate a publicly available 3DMM of the full head.

## **2.5 Facial Landmarking**

In this thesis, we use template morphing methods to establish dense correspondences. Such methods need an automatic initialisation to bring them within the convergence basin of the global minimum of alignment and morphing. Automatic initialisation is to align the input 3D data to the template, which demands 3D facial landmarks to act as sparse correspondences. The landmark detection can be done over two major data resources: 1) detection directly on the 3D mesh, 2) detection on the associated registered 2D image and project these 2D landmarks onto the 3D mesh.

Facial landmarking algorithms can be grouped into three noteworthy classes: holistic

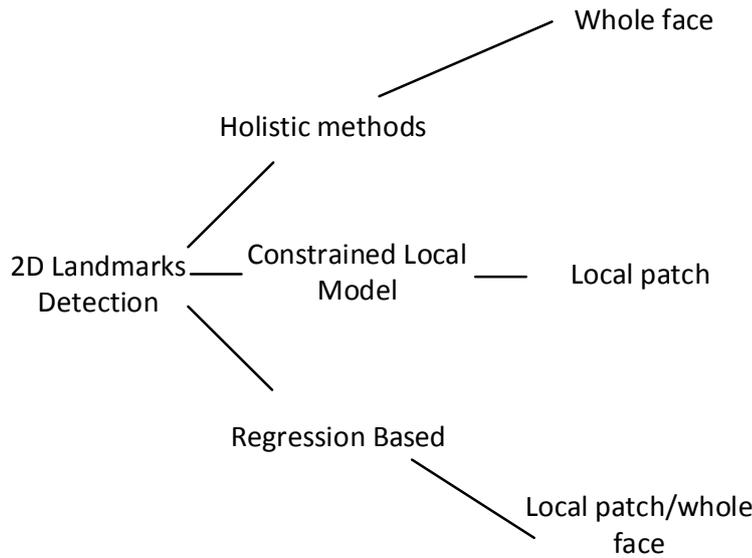


Figure 2.3: 2D facial landmarks detection systems.

methods, Constrained Local Model (CLM) methods, and regression-based methods [145]. These classes of methods may use pixel values or extracted features on the face or in local facial patches around the landmarks, as summarised in Figure 2.3. The following three subsections discuss these approaches.

### 2.5.1 Holistic Methods:

Cootes et al. [40] introduced the Active Appearance Model (AAM) which aims at fitting the facial images with a small number of coefficients using a statistical model. AMM consists of the global facial shape model and the holistic facial appearance model, which can be learned by PCA. The parameters appearance model control the shape and texture variation according to

$$x = \bar{x} + \psi_s \alpha \quad (2.1)$$

$$g = \bar{g} + \psi_t \alpha \quad (2.2)$$

where  $\bar{x}$  is the mean shape and  $\bar{g}$  is the mean texture,  $\psi_s$  and  $\psi_t$  include the shape and texture variation in the training set. The appearance model can be learned from a set of holistic pixel values or extracted features, such as SIFT [84] and HOG [49]. In order to locate the landmarks, the learned appearance and shape models are fit to the test images

for searching the best match. As a classic approach of facial landmark detection and pose estimation, many improvements over the original AAM have been proposed [74, 132]. Matthews and Baker [96] introduced an inverse compositional image alignment algorithm to solving the AAM fitting problem. This demonstrated that the appearance variation could be pre-computed and extended to use a 2D similarity transformation.

## 2.5.2 Constrained Local Models

The CLM [12, 13, 45] strategies locate facial landmarks based totally on the global facial shape models as well as the independent local appearance of each landmark. The AAM approach learns the appearance model for the whole face, while CLM models a set of local pixel values or extracted features [10]. The local features make it easier to deal with illumination variations and occlusion.

In this thesis, we make use of the Zhu and Ramanan [157] face detector and landmarker, which follows the CLM approach. Zhu and Ramanan [157] use a tree structured part model of the face, which both detects faces and locates facial landmarks. One of the major advantages of their approach is that it can handle extreme head poses even at relatively low image resolutions. A mixture of trees model is used as a part for every facial landmark. Global mixtures are subject to capturing topological adjustments because of perspective. Such mixtures for different viewpoints are shown in Figure 2.4. Zhu and Ramanan [157] demonstrated that the global mixtures can be employed to summarise the motion of deformation for a single perspective. Each tree can be represented in a linearly-parameterized and tree-structured way [148]. The score of a configuration of the parts  $L$  can be defined as:

$$S(\mathbf{I}, \mathbf{L}, \mathbf{m}) = AppM(\mathbf{I}, \mathbf{L}) + ShapeM(\mathbf{L}) + \alpha^m \quad (2.3)$$

where  $\mathbf{I}$  is the input image.  $AppM$  indicates appearance model, the evidence of which is to place a viewpoint template for part. Appearance model is built based on extracted features (e.g. HoG feature) other than directly on pixel values.  $ShapeM$  indicates shape model, the score of which is defined as mixture-specific spatial arrangement of parts.  $m$  indicates a mixture of one mixture-of-trees model.  $\alpha^m$  is a scalar bias in terms of mixture  $m$ .



Figure 2.4: Mixture-of-trees model taken from [157] encodes topological changes due to viewpoint.

### 2.5.3 Regression Methods

The regression-based strategies use machine learning methods to learn the mapping from 2D images to the landmarks [32, 53, 158]. The main difference from the previous two classes of methods is that regression methods normally do not construct any explicit global face shape constraints. The regression based techniques can be categorized into direct regression methods [144] and cascaded regression methods [54]. The former strategies predict the landmark positions in one iteration with no initialization, while the latter methods perform cascaded prediction and they normally require landmark ground-truth for learning. The regression-based methods using deep learning have a trend in prediction of 3D landmarks from 2D images [32, 53, 54].

### 2.5.4 Critical Analysis

It is less computationally expensive to detect landmarks on 2D images and project to the 3D mesh, compared to doing this directly on the 3D mesh [44]. More importantly, there are several publicly available pose-invariant 2D facial landmarks detection systems. We are aware of the recent trend in prediction of 3D landmarks from 2D images [32, 53, 54]. This can detect the repeatable facial landmarks from different pose, but it needs to solve the bias between 2D view and actual 3D locations.

## 2.6 Correspondence Establishment

Correspondence establishment is a very challenging problem regarding the fact of non-linear shape representation and non-rigid shape deformation. The main approach to dense correspondence is shape registration. As shown in Table 2.3, many methods are provided

Table 2.3: Registration methods

Transformation	Examples
Rigid Transformation	all rigid cases
Rigid Transformation with Nonrigid Deformation	[4, 34, 107]
Nonrigid Deformation	[8, 29, 30, 80, 92, 101, 102]

to register a mesh using different transformations. Four main types are common: rigid transformation, rigid deformation followed with nonrigid deformation and nonrigid deformation.

Rigid transformation follows the assumption that the objects can be registered by a Euclidean transformation consisting of rotation and translation [17, 64, 73]. This transformation, which has 6 Degrees of Freedom (DoF), is applied for all the points. When global scaling is also permitted, this is termed a similarity transformation.

Rigid transformation with non-rigid deformation allows surface to undergo a combination of global rigid transformations and local non-rigid deformation. Allen et al. [4] used a skeleton template for articulation. As related to other joints by rigid transformation, each joint allows for some DoF. Pekelny et al. [107] employed prior information of bone to estimate the transformations. Chang and Zwicker [34] used a finite set of rigid transformations to estimate the global transformations. Local deformations can be estimated by blending the transformations of the adjacent parts [34].

Nonrigid deformation methods include more generic and local deformations. There are two ways to perform deformation: displacement fields [67, 101] and local transformation [8, 92]. In the following five subsections, we describe leading non-rigid registration methods that are relevant to this thesis.

### 2.6.1 Non-rigid ICP

Amberg et al. [8] proposed the optimal-step nonrigid iterative closest point (NICP) framework, which extended Iterative Closest Point (ICP) to nonrigid deformations. Like ICP, there is an iteration loop that finds closest points between source and target shapes, which helps retain the convergence property of ICP. However, rather than the source moving

rigidly, small changes in movement are permitted between neighbouring vertices with mesh connectivity. Following Alen et al. [6] they define per-vertex locally affine deformations. Since these are under constrained, regularisation is necessary and this is achieved by minimising the weighted difference of affine transformations between neighbouring mesh vertices. The weighting is relative to other items in the cost function (primarily the sum of squared differences between corresponding vertices) and can be thought of as a mesh stiffness, which is gradually reduced through the iteration sequence. Regularised movement and deformation are termed an optimal step, which is achieved via a single linear solve within each iteration, for some level of mesh stiffness.

### 2.6.2 Global Correspondence Optimization

In contrast to rigid ICP, Li et al. [92] showed that using proximity heuristics to determine correspondences is less reliable when large deformations are present. Instead of estimating approximated corresponding points alone, the algorithm simultaneously solves for correspondences, confidence weights, and deformation field within a single non-linear least squares optimization, using the Levenberg-Marquardt algorithm. The global correspondence optimization solves simultaneously for both the deformation parameters as well as the correspondence positions.

### 2.6.3 Coherent Point Drift

Myronenko et al. consider the alignment of two point sets as a probability density estimation [100] and they call the method Coherent Point Drift (CPD). There is no closed-form solution for this optimisation, so it employs an EM algorithm to optimize the Gaussian Mixture Model (GMM) fitting. Algorithms are provided to solve for several shape deformation models such a affine (CPD-affine) and generally non-rigid (CPD-nonrigid). The ‘*non-rigid*’ motion model in [100] employs an  $M \times M$  Gaussian kernel  $\mathbf{G}$  for motion field smoothing, and the M-step requires solving for an  $M \times 3$  matrix  $\mathbf{W}$  that generates the template deformation (GMM motion field) as  $\mathbf{GW}$ . Such motion regularisation is related to motion coherence, and inspired the algorithm’s name. The approach of CPD [100] is to transform the point registration problem into a probability density estimation problem.

The template point set is represented as a Gaussian Mixture Model (GMM), where the means of the mixture components are situated at  $M$  template points,  $\mathbf{y}_m$ , in the template point set,  $\mathbf{Y}$ . The Gaussian distribution around these points is equal for all points, and isotropic with variance  $\sigma^2$ . The CPD algorithm adjusts the position of the mixture components and the variance parameter to maximise the posterior probability of the GMM model generating the  $N$  data points  $\mathbf{x}_n$  in the data point set,  $\mathbf{X}$ .

There is no closed-form solution for this optimisation and so the usual two-stage iterative procedure is adopted, the Expectation-Maximisation (EM) algorithm. In the *expectation step* (E-step), the probabilities of correspondence between template and data,  $\mathbf{P}$  are computed, initially using assumed parameters for the GMM. i.e. the initial template position and shape, and a suitable estimate of isotropic variance  $\sigma^2$ . A given entry  $P_{m,n} = P(m|\mathbf{x}_n)$  in  $\mathbf{P}$  is a posterior probability computed using Bayes theorem as:

$$P(m|\mathbf{x}_n) = \frac{\exp\left(-\frac{1}{2}\left(\frac{\|\mathbf{x}_n - \mathbf{y}_m(\theta)\|^2}{\sigma^2}\right)\right)}{\sum_{k=1}^M \exp\left(-\frac{1}{2}\left(\frac{\|\mathbf{x}_n - \mathbf{y}_k(\theta)\|^2}{\sigma^2}\right)\right) + \frac{(2\pi\sigma^2)^{\frac{3}{2}}\omega M}{(1-\omega)N}} \quad (2.4)$$

where  $\omega$  is the weighting of a uniform distribution added to the GMM to account for data noise. The symbol  $\theta$  indicates that the template points are deformed according to the current estimate of the template deformation parameters.

The *maximisation step* (M-step) then optimises the GMM parameters, based on these expected correspondence probabilities, and this two-step process iterates until convergence. Myronenko and Song [100] present three motion models for deformation of the template which they term (i) rigid, (ii) affine and (iii) non-rigid. Their ‘*rigid*’ motion model is, assuming the  $\mathbb{R}^3$  implementation, a 7-parameter similarity transform including a 6-DOF 3D rigid transform and a global scale parameter, whereas the affine motion model is the standard unconstrained 12 parameter formulation. For the affine case they optimise for

$$\mathbf{B} = f_B(\mathbf{X}, \mathbf{Y}, \mathbf{P}), \mathbf{t} = f_t(\mathbf{X}, \mathbf{Y}, \mathbf{P}), \sigma^2 = f_\sigma(\mathbf{X}, \mathbf{Y}, \mathbf{P}) \quad (2.5)$$

where  $\mathbf{B}$  is an affine transformation matrix (excluding translation),  $\mathbf{t}$  is a translation vector and  $\sigma^2$  is a scalar representing variance. We omit the functional forms of  $f_B, f_t, f_\sigma$  given in [100] due to page restrictions.

The CPD method has been extended by various groups [70, 78, 133, 137]. Compared to TPS-RPM, CPD offers superior accuracy and stability with respect to non-rigid deformations in presence of outliers. A modified version of CPD imposed a *Local Linear Embedding* topological constraint to cope with highly articulated non-rigid deformations [65]. However, this extension is more sensitive to noise than CPD. A non-rigid registration method used Student Mixture Model (SMM) to do probability density estimation [156]. The results are more robust and accurate on noisy data than CPD. Our morphing technique (Dai et al. [47]), a hierarchical parts-based CPD-LB morphing framework, is able to avoid under-fitting and over-fitting. It overcomes the sliding problem to some extent, but the end result still has a small tangential error.

Recent techniques emphasized intrinsic models and performed isometric deformation. Intrinsic geometry concerns geometry properties, such as face area and angle. Jain et al. [80] sought a low-dimensional embedding method to preserve all pairwise geodesic distances. [29] used generalized multidimensional scaling to embed one mesh in another for partial mesh matching. Based on previous work, Bronstein et al. [30] used a different distance measurement known as diffusion distance and Gromov-Hausdorff distance to deal with topological noise. The result in [102] demonstrated that a single correspondence was able to be established for all points by means of the heat kernel.

#### 2.6.4 Laplace-Beltrami Mesh Manipulation

The Laplace-Beltrami (LB) operator is widely used in 3D mesh manipulation. The LB term regularises the mesh manipulation in two ways: 1) the manipulated points on the mesh template are forced to move towards their corresponding position on the raw 3D scan; 2) all other points in mesh template are moved as rigidly as possible regarding the manipulated points' movement, according to an optimised cost function.

Following Sorkine et al. [121], the idea for quantifying the rigid deformation energy is to sum up the deviations from rigidity. Thus, the energy functional can be formed as:

$$\mathbf{E}(\mathbf{S}') = \sum_{i=1}^n \mathbf{w}_i \sum_{j \in N(i)} \mathbf{w}_{ij} \|(\mathbf{p}'_i - \mathbf{p}'_j) - \mathbf{R}_i(\mathbf{p}_i - \mathbf{p}_j)\|, \quad (2.6)$$

where we denote a mesh by  $\mathbf{S}$ , with  $\mathbf{S}'$  its deformed mesh and  $\mathbf{R}$  is a rotation. The mesh consists of  $n$  vertices and  $m$  faces. Also  $N(i)$  is the set of vertices connected to vertex

*i*. The parameters  $w_i, w_{ij}$  are fixed cell and edge weights. Note that  $E(S')$  relies on the geometries of  $S, S'$ , which is also on the vertex positions  $\mathbf{p}, \mathbf{p}'$ . Specifically, since the reference mesh (our input shape) is fixed, the only variables in  $E(S')$  are the deformed vertex positions  $\mathbf{p}'_i$ . The gradient of  $E(S')$  is computed with respect to the positions  $\mathbf{p}'$ . The partial derivatives  $\mathbf{p}'_i$  can be written as:

$$\frac{dE(S')}{d\mathbf{p}'_i} = \sum_{j \in N(i)} 4w_{ij} \left( (\mathbf{p}'_i - \mathbf{p}'_j) - \frac{1}{2}(\mathbf{R}_i + \mathbf{R}_j)(\mathbf{p}_i - \mathbf{p}_j) \right) \quad (2.7)$$

Setting the partial derivatives to zero, each  $\mathbf{p}'_i$  represents the following linear system of equations:

$$\sum_{j \in N(i)} w_{ij}(\mathbf{p}'_i - \mathbf{p}'_j) = \sum_{j \in N(i)} \frac{w_{ij}}{2}(\mathbf{R}_i + \mathbf{R}_j)(\mathbf{p}_i - \mathbf{p}_j) \quad (2.8)$$

The linear combination on the left-hand side is the discrete Laplace-Beltrami operator applied to  $\mathbf{p}'$ , hence the system of equations can be written as:

$$\mathbf{L}\mathbf{p}' = \mathbf{b}, \quad (2.9)$$

where  $\mathbf{b}$  is an  $n$ -vector whose  $i$ -th row contains the right-hand side expression from (2.8). We also need to incorporate the modeling constraints into this system. In the simplest form, those can be expressed by some fixed positions

$$\mathbf{p}'_j = \mathbf{c}_k, k \in \mathcal{F}, \quad (2.10)$$

where  $\mathcal{F}$  is the set of indices of the constrained vertices.

### 2.6.5 Parametrisation Methods

Parametrisation is another approach to shape correspondence. Parametrisation of shape allows points insertion, points sliding, and points deletion to establish correspondence. Some methods [33] use geodesic distances between each landmark to insert corresponding points. Others tend to find the best parametrisation among all the subjects [50].

**Contour Subdivision:** Correspondences can be found by localisation of salient feature points ('landmarks') and then by dividing the length of the contour equally in terms of geodesic distances between each landmark. Landmarks are detected and labelled by virtue of their learned local properties and their relative configuration. Relative local

properties can be enforced by a model-fitting approach based on sample consensus, such as RANSAC [33].

**Minimum Description Length:** Davies et al. [50] posed the correspondence establishment as one of finding the best parameterisation for each shape in the training set. The algorithm selects the set of parameterisations by minimising the description length of the training set. This objective function captures both the model complexity itself and the data values required to express each of the training scans with the model.

### 2.6.6 Critical Analysis

There exists lots of correspondence establishment methods including template morphing, isometric deformation and shape parameterisation. The most recent progress in 3DMM construction pipelines reveals that template morphing method is a better choice in this case than other methods [124]. The goal of template morphing is to align and morph a *source* point set (or mesh) to a *target* point set. By using some form of template shape as the source, non-rigid shape morphing is able to reparametrise a collection of raw 3D scans of some object class into a consistent form. This is feasible when the optimisation of a cost function forces the template to morph towards the shape of target mesh. The cost function can be defined as a distance error or a functional energy between the morphed template and the target mesh. The most recent progress has a trend in correspondence matching by deep learning [28, 119, 142]. Once trained, correspondence matching can be a real-time process. Lots of features in the face can be used for the optimisation of the cost function. However, this is not the case in cranium, which makes it a more complicated problem than face-only template morphing.

With correct correspondences, a 3DMM is able to sensibly describe the shape variation within the object class [123]. The ground-truth of dense correspondence is required for the evaluation of correspondence accuracy. When the ground-truth of dense correspondence is often unknown, the manual landmarks can be used as the ground-truth in evaluation [123, 124]. Based on this evaluation metric, Ericsson et al. [61] used the mean absolute distance, which is calculated between the manual landmarks and points corresponding to the same landmarks in a mesh template, for benchmarking.

## 2.7 Procrustes Alignment

In some techniques in the literature, establishing correspondences and attaining alignment are tightly coupled, whereas in others they are more separate. In previous section, we consider the tightly coupled methods and in following section, we consider the standalone alignment procedure GPA [71, 72, 126], which operates on the assumption that correspondences are known. It is of course possible to realign, if there is good rationale to do so, using GPA after correspondences have been found using any method.

We now briefly recap the well-known Procrustes alignment procedure. The collection of scans in dense correspondence are subjected to Generalised Procrustes Analysis (GPA) to remove similarity effects (rotation, translation, scale), leaving only shape information. Let  $X$  and  $Y$  be a pair of  $(m \times n)$  matrices representing two objects with  $n$  corresponding points in  $m$  dimensional space. Full Procrustes alignment seeks an optimal similarity transformation with  $(\mathbf{R}, \mathbf{t}, \mathbf{s})$  denoted as rotation, translation and scaling respectively, such that

$$(\mathbf{R}, \mathbf{t}, \mathbf{s}) = \arg \min_{\mathbf{R}, \mathbf{t}, \mathbf{s}} \|\mathbf{s}(\mathbf{R}\mathbf{X} + \mathbf{t}\mathbf{u}^T) - \mathbf{Y}\|_2^2 \quad (2.11)$$

where  $\mathbf{u}$  is an  $n$ -dimensional column vector of unit values. If one is interested in form models, the scaling is unity, otherwise it is determined by ensuring  $X$  and  $Y$  have the same mean (or RMS) distance to centroid. The translation component can be eliminated by centering all shapes on their centroids. The solution to the rotation can readily be found using Singular Value Decomposition (SVD).

Generalising the above to many sets of corresponding points, Generalized Procrustes Analysis (GPA) is a widely used alignment method in statistical modeling for aligning training datasets. It aims to align to the mean shape of the dataset, but the mean shape depends on alignment and so the GPA process is iterative, as follows:

1. Set the initial estimate of the mean shape to one example from the data set (e.g. the first or the nearest to mean size);
2. Align all the remaining shapes to the estimated mean shape (updated in every loop);
3. If the estimated mean has changed significantly, go to step 2.

The convergence of GPA alignment is detected by the stabilization of the mean shape.

### 2.7.1 Critical Analysis

A collection of scans in dense correspondence can be subjected to Procrustes alignment to remove similarity effects (rotation, translation, scale), leaving only shape information where GPA is the standard method. Since the size of human face grows with age, it is of great interests to model this motion of face growth. So it is still questionable to remove scaling among the similarity effects when doing Procrustes analysis. In our case, we suggest that we should retain the size information in constructing 3DMMs of the face/head.

## 2.8 Statistical Modelling

Statistical modelling is the final step of 3DMM construction. It represents shape using a reduced set of parameters and models their multivariate distribution. The number of parameters is small relative to  $3N$  where  $N$  is the number of 3D points in the mesh. PCA is one form of this that decorrelates the parameters that control shape variation. The standard method (PCA) and recent progress are discussed in the following sections.

### 2.8.1 Principal Component Analysis

The processed meshes are statistically analysed, typically with Principal Component Analysis (PCA), generating a 3DMM as a linear basis of shapes. This allows for the generation of novel shape instances. Each deformed template is represented by  $p$  3D points  $(x_i, y_i, z_i)$  and is reshaped to a  $3p$  row vector. Each of these vectors is then stacked in a  $k \times 3p$  data matrix, and each column is made zero mean. Singular Value Decomposition is applied from which eigenvectors are given directly and eigenvalues can be computed from singular values. This yields a linear model as:

$$\mathbf{X}^* = \bar{\mathbf{X}} + \mathbf{P}\mathbf{U} = \bar{\mathbf{X}} + \sum_{i=1}^k \mathbf{P}_i U_i \quad (2.12)$$

where  $\bar{\mathbf{X}}$  is the mean head shape vector and  $\mathbf{P}$  is a matrix whose columns  $\mathbf{U} = [U_1, \dots, U_k]$  are the eigenvectors of the covariance matrix. The vector  $\mathbf{P}$  holds the shape parameters

$\{P_i\}$ , that weight the shape variation modes which, when added to the mean shape, model a shape instance  $\mathbf{X}^*$ .

## 2.8.2 Gaussian Process Morphable Model

Marcel et al. [94] employed Gaussian process (GP) to statistically model the shape variations. The modelling by GP is represented by the leading components of its Karhunen-Loeve expansion. A Gaussian Process Morphable Model (GPMM) uses manually defined arbitrary kernel functions to describe the deformation's covariance matrix. This enables a GPMM to aid the construction of a 3DMM, without the need for training data. The posterior models (PMs) of GPMMs are regression models of the shape deformation field. Given partial observations, such posterior models are able to determine what is the potential complete shape. A posterior model is able to estimate other points' movements when some set of landmarks and their target positions are given.

A Gaussian Process Morphable Model (GPMM) uses manually defined arbitrary kernel functions to describe the deformation's covariance matrix. This enables a GPMM to aid the construction of a 3DMM, without the need for training data. The posterior models of GPMMs are regression models of the shape deformation field. Given partial observations, such posterior models are able to determine what is the potential complete shape. A posterior model is able to estimate other points' movements when some set of landmarks and their target positions are given.

Instead of modelling absolute vertex positions using PCA, GPMMs represent the shape variation as a vector field  $\mathbf{u}$  from a template  $\mathbf{X} \in \mathbb{R}^{p \times 3}$  towards a target shape  $\mathbf{X}'$  represented as

$$\mathbf{X}' = \mathbf{X} + \mathbf{u}(\mathbf{X}) \quad (2.13)$$

for some deformation vector field  $\mathbf{u} \in \mathbb{R}^{p \times 3}$ . We model the deformation as a Gaussian process  $\mathbf{u} \sim GP(\mu, \mathbf{k})$  where  $\mu \in \mathbb{R}^{p \times 3}$  is a mean deformation and  $\mathbf{k} \in \mathbb{R}^{3 \times 3}$  a covariance function or kernel. The core idea behind this approach is that a parametric, low-dimensional model can be obtained by representing the Gaussian process using the  $r$  leading basis functions  $\psi_i \in \mathbb{R}^{p \times 3}$  of its Karhunen-Loève expansion:

$$\mathbf{u} = \mu + \sum_{i=1}^r \alpha_i \sqrt{\lambda_i} \psi_i, \alpha_i \in \mathcal{N}(0, 1) \quad (2.14)$$

Since GPMMs have much more freedom in defining the covariance function, much more shape variation is modelled when compared to standard PCA models. However, the shape generated by such models may not well-represent the shape required.

### **2.8.3 Statistical Modelling using Auto-encoder**

Recently Ranjan et al. [111] introduced a Convolutional Mesh Autoencoder (CoMa) consisting of mesh downsampling and mesh upsampling layers with fast localized convolutional filters defined on the mesh surface. Bagautdinov et al. [11] proposed a method to model multi-scale face geometry that learns the facial geometry using UV parameterization for mesh representation, which started with the observation that both global and local linear models can be viewed as specific instances of autoencoders. This approach featured a variational autoencoder with multiple layers of hidden variables that capture various level of geometrical details.

### **2.8.4 Critical Analysis**

PCA is a very well known and widely used procedure that works well for linear shape spaces. Gaussian processes and auto-encoders can replace PCA in statistical modelling. Deep learning methods may be used methods, but are not the focus of the thesis.

## **2.9 3D Facial Symmetry and Asymmetry**

Thesis contributions include symmetric morphing and explicit modelling of symmetric and asymmetric shape variation. Most biological objects, including human heads and bodies, possess approximate symmetries. Often this is principally extrinsic, bilateral symmetry (i.e. reflective symmetry about a vertical plane bisecting the object). Deviations from exact symmetry are an interesting and potentially important geometric property in terms of modelling and understanding 3D shape variation. The degree of asymmetry may convey information about an object. For example, there is evidence that facial asymmetry is used by humans to measure genetic health [81] and asymmetry in man-made objects may indicate imperfections in the manufacturing process.

There is a large body of work on detecting symmetries in images [104], 2D shapes, 3D meshes and point clouds. These symmetries could be exact or approximate [98], extrinsic [110] or intrinsic [103], partial [104, 118], hierarchical or full [103]. Once detected, this enables symmetry-aware mesh processing [68] to take place in which symmetries help regularise and denoise a range of processing tasks.

The geometric morphometrics community [87, 89, 90, 95] have built models of deviations from symmetry, though this has largely been in 2D. Savriama et al. [115] present a decomposition of asymmetric shape into a symmetric shape and asymmetry variation. There are many works on how to measure facial asymmetry variation. There is much literature from both the Computer Vision perspective [35, 93, 97, 159] and the Biology perspective [38, 75, 89].

The ICP based approach to computing the symmetry plane of bilateral objects in point sets [51, 143, 153] employed a rigid-body transformation-based approach to estimate the symmetry plane. The rigid-body transformation called ICP algorithm is an iterative rigid shape alignment algorithm developed by Besl et al. [18]. Given some shape data, with any shape structures that we wish to align the data to. We can refer to points only and thus ICP can be summarised in three discrete iterative steps, as follows: 1) Compute the closest points; 2) Compute the aligning rigid transformation; 3) Apply the aligning rigid transformation. The major advantage of this approach is to benefit from the ICP algorithm relying on known closed-form solutions for the absolute orientation problem [9, 63] to find transformation matrix. However, computing the symmetry plane from the optimal rigid-body transformation is an ill-posed problem and has different solutions.

The concern about facial asymmetry analysis is that most of the existing symmetry plane estimation methods took no consideration of accurate correspondence between the shape and its reflection. The closest point correspondence searching method is commonly used in the symmetry plane estimation methods. However, the closest point is obviously not the best choice as there are a lot of more accurate correspondence methods [8, 92, 100].

### **2.9.1 Critical Analysis**

Craniofacial symmetry and asymmetry are very interesting properties to exploit in template morphing and statistical modelling. Craniofacial symmetry can be used as a corre-

spondence constraint, which has the potential to improve the accuracy of correspondence. Statistical shape models should separate symmetric from asymmetric shape variability in order to reveal interesting and potentially subtle aspects of shape variation over population datasets.

## 2.10 Existing 3DMM Construction Pipelines

A 3DMM construction pipeline is an automatic procedure that includes initialisation, dense correspondence establishment, alignment and statistical modelling methods. The input is a 3D face dataset and the output is a 3D morphable model. The pipeline should be able to be demonstrated on a publicly-available dataset. We compare the two recent works of 3DMM construction pipeline: the LSFM pipeline (2017) [27], and the Basel' Open Framework (2017) [67]. Later in the thesis, we compare these pipelines with the one that we propose, which is called the LYHM pipeline.

### 2.10.1 LSFM Pipeline

LSFM used a 2D facial landmarker to detect 2D facial landmarks and then project them onto 3D mesh. The 3D landmarks are used to globally align all meshes to a face template mesh. With such a large cohort of data, there will be some convergence from either landmarking error or NICP. They employed an error pruning process to avoid undesirable deformation results because of the failures of algorithm convergence. Then the registered data is processed by GPA for alignment, and PCA for the shape variation modelling.

### 2.10.2 Basel Open Framework

Basel Open Framework (OF) [67] requires manual landmarks for the initialisation of registration. They employed GPMM registration to establish dense correspondence, where the model is obtained by modelling the possible deformation of a reference mesh, using a Gaussian process. This initial GPMM models not the actual shape variation but the possible deformation defined by a combination of handcrafted Gaussian kernels. They have five level regularisation terms corresponding to five parts in the face. The deformation

Table 2.4: Pipelines comparison

	Initialisation	Dense correspondence	Alignment	Modelling
LSFM	Automatic facial landmarks	NICP with error pruning	GPA	PCA
OF	Manual landmarks needed	GPMM registration	GPA	GP

of ear is the most regularised and the nose is the least regularised. There usually exists much noise in the ear region of 3D mesh. More regularisation makes the ear deformation more robust to the noise. In order to capture the detail of local region, less regularisation enables more flexible deformation. Then the registered data is subjected to GPA to remove similarity effects. The shape variation is modeled by Gaussian process using the leading components of its Karhunen-Loeve expansion. The 3DMM construction pipeline comparison can be seen in Table 2.4.

### 2.10.3 Critical Analysis

Well-known 3DMMs of the human face are the BFM and LSFM, which also present their own facial 3DMM construction pipelines. The LSFM pipeline used the same NICP template morphing approach as the BFM pipeline, but improved by an error pruning method. The new version of the BFM has adopted the GPMM registration framework for correspondence establishment and the Gaussian process for statistical modelling. They are widely used in 3D face reconstruction from 2D images. LSFM uses its automatic landmarking system on synthetic images. The OF needs manual landmarks, which is highly time-consuming. We aim to enable fully-automatic 3DMM construction of the full human head and we compare the LSFM and OF pipelines with our proposed pipeline in Chapter 3.

## 2.11 Evaluation Criteria

For quantitative evaluation of the 3DMM construction pipelines, three performance metrics are used for evaluation, compactness, generalization and specificity [123].

### 2.11.1 Compactness

The compactness of the statistical shape model describes the number of parameters required to express some fraction of the variance in the training set used to construct that model. A more compact model used fewer parameters to express a given fraction of variance, or expresses a bigger fraction of variance for a given number of shape parameters. So compactness measures the efficiency of a model to capture the shape variability. Compactness can be defined as:

$$C(m) = \frac{\sum_{i=1}^m \alpha_i}{\sum_{i=1}^N \alpha_i}. \quad (2.15)$$

where  $\alpha_i$  is the  $i$ -th eigenvalue,  $m$  is the number of shape parameters and  $N$  is the total number of shape parameters in the model.

### 2.11.2 Generalization

Generalization measures the capability of the model to represent unseen examples of the class of objects. It can be measured using the *leave-one-out* strategy, where one example is omitted from the training set and used for reconstruction testing. The accuracy of describing the unseen example is calculated by the mean point to point Euclidean distance error. With an increasing number of model parameters, the generalization error is expected to decrease. For the same number of model coefficients, the lower mean Euclidean distance error, the better the model.

### 2.11.3 Specificity

Specificity measures the ability to generate shape instances of the class that are similar to those in the training set. In order to assess specificity, a set of shape instances should be sampled from the shape space. Then the Euclidean distance error to the closest training shape is calculated for each shape instance and the average is taken for all the shape instances. The mean Euclidean distance error is expected to increase with increasing number of parameters, as the increasing number of PCA coefficients gives more flexibility to shape reconstruction. It also increase the likelihood of the reconstructed shape instances being away from the real data. For specificity measurement, the lower Euclidean distance error, the closer the model is to the training data data, so the specificity is better.

### 2.11.4 Critical Analysis

A high quality of the correspondences is desirable to improve the performance of 3DMM. The compactness of the model describes the number of shape components required to express some fraction of the variance in the training set. More shape components enable the 3DMM to retain more shape variation in the dataset. The generalisation of the model demonstrates the ability in sketching unseen examples. The combination of more shape components has better performance in describing the unseen example. Specificity measures how well a model is able to generate instances that are similar to real data. In contrast to the previous two metrics, more shape components includes more noisy shape variation [123], which is not desirable for generating instances that are similar to real data. Overall, more components is not always desirable for the performance of 3DMM. There is a trade-off among compactness, generalisation and specificity.

## 2.12 Summary

The last 20 years of research in 3DMM construction has mainly focused on two main techniques: 1) dense correspondence establishment and 2) statistical modelling methods. The theme throughout this thesis is to develop better correspondence establishment algorithms and statistical modelling methods.

For dense correspondence establishment, one promising method is template morphing algorithms which is to find the best combination of global and local transformation. This is feasible when optimising a cost function that forces the template to morph towards the shape of target mesh. The cost function can be defined as distance error or functional energy between morphed template and target mesh. The most recent progress has a trend in correspondence matching by deep learning.

In terms of statistical modelling methods, the standard method is PCA that finds the directions in the vector space that have maximum variance, whilst being mutually orthogonal. Multi-linear models use several PCA models built from sub-regions of the meshes with a post process merging them into a global PCA model. A Gaussian process can model shape variation by just defining different Gaussian kernels. It still loses high frequency signals in statistical modelling. Recently, there exists a trend in statistical

modelling for using auto-encoders. The input meshes either need re-meshing before the learning process when processing 3D meshes directly, or need to be transferred into a functional space, e.g. UV map representation. At the end of the decoder, there requires a back-transformation from the functional space to 3D spatial coordinates.

## Chapter 3

# Modelling of Orthogonal Craniofacial Profiles

In the medical analysis of craniofacial shape, the visualisation of 2D profiles [112] is highly informative when looking for deviations from population norms. It is often useful, in terms of visual clarity and attention focus, for the clinician to examine shape outlines from canonical viewpoints; for example, pre- and post-operative canonical profiles can be overlaid. We view profile-based modelling and analyses as being complementary to that of a full 3D shape model. Profile visualisations should be backed up by quantitative analysis, such as the distance (in standard deviations) of a patient's shape profile from the mean profile of a reference population. Therefore, we have developed a novel image processing pipeline to generate a 2D morphable model of craniofacial profiles from a set of 3D head surface images. Building on the initial work of Pears and Duncan [106], who modelled a single craniofacial profile over the sagittal plane, we construct morphable 2D profile models over three orthogonal planes to provide comprehensive models and analyses of shape outline.

The chapter is structured as follows. Section 3.1 discusses our new pipeline used to extract profiles and construct 2D morphable models. Section 3.2 evaluates several variants of the constructed models both qualitatively and quantitatively and compares our single-profile models with our multi-profile model. A final section concludes the work.

### 3.1 Model Construction Pipeline

Our pipeline to build a 2D morphable model is illustrated in Figure 3.1. Input data (left of figure) consists of 1212 3D images of subjects with an even split of males and females. This is a subset of the Headspace dataset, captured using 3dMD's 5-camera 3dMD head system. All subjects are wearing tight-fitting latex caps to expose the shape of the cranium, and we excluded around 300 scans from the dataset, primarily due to visible hair bulges or poor fitting of the cap. The processing pipeline in Figure 3.1 employs a range of techniques in both 3D surface image analysis and 2D image analysis and has three main stages:

- (i) 2D shape extraction: The raw 3D scan from the Headspace dataset undergoes pose normalization and pre-processing to remove redundant data (lower neck and shoulder area), and the 2D profile shape is extracted as closed contours from three orthogonal viewpoints: the side view, top view and frontal view (note that we automatically remove the ears in the top and frontal views, as it is difficult to get good correspondences over this section of the profiles).
- (ii) Dense correspondence establishment: A collection of profiles from a given viewpoint is reparametrised into a form where each profile has the same number of points joined into a connectivity that is shared across all profiles.
- (iii) Similarity alignment and statistical modelling: The collection of profiles in dense correspondence are subjected to Generalised Procrustes Analysis (GPA) to remove similarity effects (rotation, translation and scale), leaving only shape information. The processed meshes are statistically analysed, typically with PCA, generating a 2D morphable model expressed using a linear basis of eigen shapes. This allows for the generation of novel shape instances, over any of the three viewpoints.

Each of the profiles is represented by  $m$  2D points  $(y_i, z_i)$  and is reshaped to a  $2m$  row vector. Each of these vectors is then stacked in a  $n \times 2m$  data matrix, and each column is made zero mean. Singular Value Decomposition (SVD) is applied from which eigenvectors are given directly and eigenvalues can be computed from singular values.

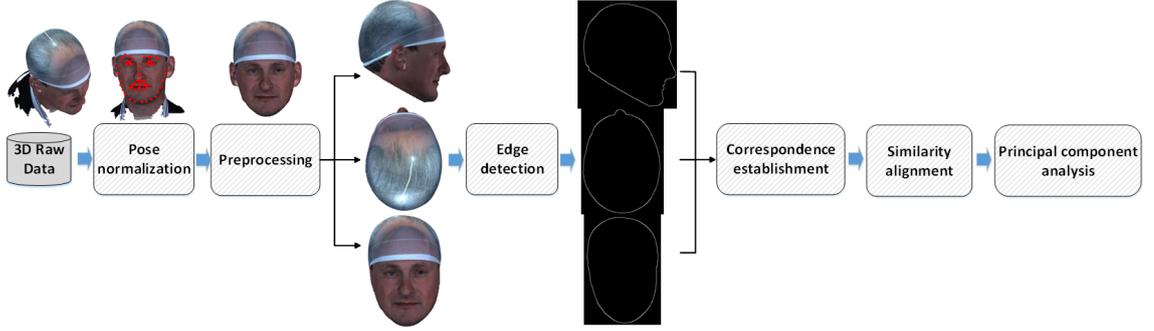


Figure 3.1: The pipeline for 2D morphable model construction. Note that 2D profiles are extracted as closed contours from three orthogonal viewpoints.

This yields a linear model as:

$$\mathbf{x}_i = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}_i = \bar{\mathbf{x}} + \sum_{i=1}^k \mathbf{p}^k b_i^k \quad (3.1)$$

where  $\bar{\mathbf{x}}$  is the mean head profile shape vector and  $\mathbf{P}$  is a matrix whose columns  $\mathbf{p}^k$  are the eigenvectors of the covariance matrix (after pose alignment), describing orthogonal modes of head profile variation. The vector  $\mathbf{b}$  holds the shape parameters  $\{b^k\}$  that weight the shape variation modes, which when added to the mean shape, model a shape instance  $\mathbf{x}_i$ . The three main stages of the pipeline are described in the following subsections.

### 3.1.1 2D Shape Extraction

2D shape extraction requires three stages, namely (i) pose normalisation, (ii) cropping, (iii) edge detection and (iv) subdivision. Each of these stages is described in the following subsection.

**Pose Normalisation:** Using the colour-texture information associated with the 3D mesh, we can generate a realistic 2D synthetic image from any view angle. We rotate the scan over 360 degrees in pitch and yaw (10 steps of each) to generate 100 images. Then, the Viola–Jones face detection algorithm [135] is used to find the frontal face image among this image sequence. A score is computed that indicates how frontal the pose is. The 2D image with the highest score is chosen to undergo 2D facial landmarking. We employ the method of Constrained Local Models (CLMs) using robust discriminative response map fitting [10] to do the 2D facial image landmarking. Then, the trained system

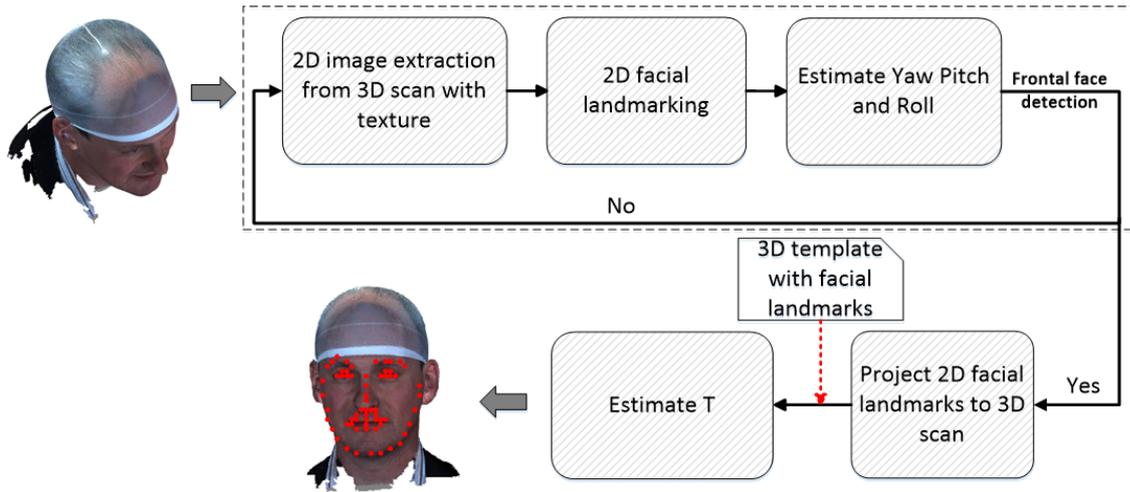


Figure 3.2: 3D pose normalization using the texture information.

is used to estimate the three angles for the image with facial landmarks. Finally, 3D facial landmarks are captured by projecting the 2D facial landmarks to the 3D scan. As shown in Figure 3.2, by estimating the rigid transformation matrix  $T$  from the landmarks of a 3D scan to that of a template, a small adjustment of pose normalization is implemented by transforming the 3D scan using  $T$ .

**Cropping:** 3D facial landmarks can be used to crop out redundant points, such as the shoulder area and long hair. The face landmarks delineate the face size and its lower bounds on the pose normalised scan, allowing any of several cropping heuristics to be used. We calculate the face size by computing the average distance from facial landmarks to their centroid. Subsequently, a plane for cropping the 3D scan is generated by moving the cropping plane downward an empirical percentage of the face size. We use a sloping cropping plane so that the chin area is included, but that still allows us to crop close to the base of the latex skull cap at the back of the neck to remove the (typically noisy) scan region, where the subject's hair emerges from under the cap (see Figure 3.1).

**Edge Detection:** We use side view, top view and frontal view from the 3D scan to reveal the 2D profile shape, and we can generate a 2D contour within the three views by orthogonal projection. For example, in the side view (Y-Z view), we traverse the Y direction in small steps, and at each step, we compute the minimum and maximum Z value. The points with the minimum and maximum Z value are the contour points in the side view.

**Subdivision:** In order to extract profile points using subdivision, an interpolation procedure is needed to ensure there is a fixed number of profile points across all profiles. The cranial region is smooth and approximately elliptical in structure. In particular for the cranial region, vectors are projected from the ellipse centre. Then we intersect a set of fitted cubic spline curves with a fixed angle, which starts from the nasion.

### 3.1.2 Correspondence Establishment

As well as using subdivision points directly in model construction, we form a model template as the mean of the population of subdivided and aligned profiles, and we use template deformation on the dataset. The resulting deformed templates are re-parametrised versions of each subject that are in correspondence with one another. The correspondence establishment including two steps:

1. use a machine learning method to find facial landmark localisations.
2. use the automatic landmarks for rigid alignment to bring all profiles within the convergence basin of the global minimum of alignment and morphing.
3. employ template morphing algorithms to reparametrise all the profiles into a consistent form where the number of vertices, the triangulation, and the (approximate) anatomical meaning of each points.

**Automatic Annotation** Following the work of Pears and Duncan [106], we employ a machine learning method to find facial landmark positions. In this method, a disc is centered on some point on the head profile among the largest scale that we are interested in. Then we fit a quartic polynomial to the profile points within that disc, which can be solved by least squares. We can also select a quartic to fit to the area on the lips with flexibility. Thus, quartic parameters  $p^T$  can be solved to fit a set of profile points  $[\mathbf{x}_p, \mathbf{y}_p]$  such that, with  $n = 4$ :

$$\hat{y}_p = \mathbf{p}^T \mathbf{x}_p, \quad \mathbf{p} = [p_0 \dots p_n]^T, \quad \mathbf{x}_p = [x_p^0 \dots x_p^n]^T \quad (3.2)$$

In order to implement the disc operator, a dense set of  $n$  point samples are selected within that disc,  $[\mathbf{x}_d, \mathbf{y}_d]$  and the operator value can be computed as:

$$\alpha = \frac{1}{n} \sum_{i=1}^n \text{sign}(y_d - \mathbf{p}^T \mathbf{x}_d) \quad (3.3)$$

When the values of  $(-1 \leq \alpha \leq 1)$  are close to zero, this indicates locally flat regions. The positive values indicates convexities, such as nose tip, while negative values indicates concavities, such as the nasion. This pattern is a discrete approximation to finding the area, which gives a high frequency noise on the signal. This can be filtered with a 10th order low pass Butterworth filter.

**Rigid Alignment:** We then use the automatic landmarks for rigid alignment from profile points to a morphing template. The same landmarks are picked once on the morphing template  $[\mathbf{x}_t, \mathbf{y}_t]$ . Given the indices of automatic landmarks  $M$  from profile points  $[\mathbf{x}_p, \mathbf{y}_p]$  and the indices of the same landmarks  $N$  from the morphing template  $[\mathbf{x}_t, \mathbf{y}_t]$ , the rigid alignment transformation matrix can be solved from:

$$(\mathbf{R}, \mathbf{t}) \rightarrow \underset{\mathbf{R}, \mathbf{t}}{\operatorname{argmin}} \|\mathbf{x}_t(N), \mathbf{y}_t(N) - ([\mathbf{x}_p(M), \mathbf{y}_p(M)]\mathbf{R} + \mathbf{t})\|^2 \quad (3.4)$$

where  $\mathbf{R}$  is the rotation matrix and  $\mathbf{t}$  is the transformation matrix. This can be solved by ICP algorithm.

**Template Morphing:** We form a model template as the mean of the population of subdivided and aligned profiles, and we use template deformation on the dataset. In this paper, we apply subdivision, Thin Plate Splines (TPS) [24], Non-rigid Iterative Closest Points (NICP) [8], Li's method [92], Coherent Point Drift (CPD) [100] and Minimum Description Length (MDL) [50] to the proposed pipeline for comparative performance evaluation.

### 3.1.3 Similarity Alignment

We use similarity alignment methods to remove the similarity effects such as rotation and translation. We use both the standard GPA approach and Ellipse Centre Nasion (ECN) method [106]. We compare the performance of these two alignment methods along with different correspondence establishment methods. The ECN method was motivated by the fact that the cranial shape appears to be elliptical. This suggests a natural centre and frame origin when aligning craniofacial profiles. The large shape variation in the nose induces displacements in the cranial alignment, which causes wrong alignment in the cranium. So in ECN method the nasion's position is used to segment out the cranium region from the face region.

## 3.2 Morphable Model Evaluation

We built four 2DMM variants of the side-view profile from the Headspace dataset and animated shape variation along the principal components (later, the full dataset is used). The four model variations correspond to full head, scale normalised and unscaled, and cranium only, scale normalised and unscaled.

For quantitative evaluation of morphable models, Styner et al. [123] give detailed descriptions of three metrics: compactness, generalisation and specificity, now used on our scale-normalised models.

**Compactness:** This describes the number of parameters (fewer is better) required to express some fraction of the variance in the training set. As illustrated in Figure 3.3, the compactness using ECN alignment is superior to that of GPA alignment, for all correspondence methods. Among these correspondence methods, subdivision, TPS and MDL, all aligned with ECN, are able to generate the most compact models.

**Specificity:** Specificity measures the model’s ability to generate shape instances of the class that are similar to those in the training set. In order to assess specificity, a set of shape instances should be sampled from the shape space. Then the Euclidean distance error to the closest training shape is calculated for each shape instance and the average is taken for all the shape instances. We generate 1000 random samples and take the average Euclidean distance error to the closest training shape for evaluation; lower is better. We show the specificity error as a function of the number of parameters in Figure 3.4. Across all correspondence methods with GPA, it gives better specificity against all correspondence methods with ECN. This suggests that GPA helps improve the performance of modelling the underlying shape space. NICP with GPA capture the best specificity.

**Generalisation:** Generalisation measures the capability of the model to represent unseen examples of the class of objects. It can be measured using the leave-one-out strategy, where one example is omitted from the training set and used for reconstruction testing. The accuracy of describing the unseen example is calculated by the mean point-to-point Euclidean distance error; the lower the better. Generalization results are shown in Figure 3.4, and for more parameters, the error decreases, as expected. NICP with GPA performs better in terms of Euclidean distance once less than seven model dimensions are used. Between seven and 20 model dimensions, TPS with ECN outperforms other methods. When

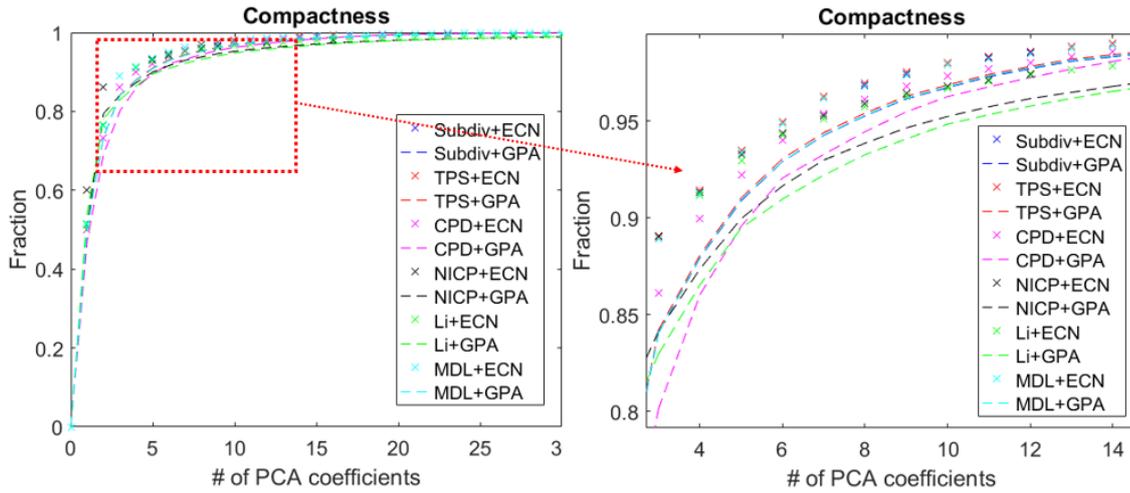


Figure 3.3: Compactness; higher is better. The right figure is used to zoom into where the differences between the methods is greatest. Alignment methods: ECN, Ellipse Centre Nasion; GPA, Generalised Procrustes Analysis. Correspondence methods: Subdiv, Subdivision; TPS, Thin Plate Splines; NICIP, Non-rigid Iterative Closest Points; Li, Li’s method [92]; CPD, Coherent Point Drift; MDL, Minimum Description Length. Please view in colour to distinguish the plots.

more than 20 model dimensions are used, CPD with GPA has the best generalization ability. Overall, GPA is able to help more successfully model the underlying shape against ECN for the same correspondence method, thereby generating better reconstructions of unseen examples.

### 3.2.1 Single Profile vs. Global Multi Profiles

We used the proposed pipeline to build three separate viewpoint 2DMM profile models with GPA and PCA with over 1212 subjects and an even split of males and females. The variations of the four main modes are shown in Figure 3.5. These models can be used for the analysis of 2D shape from the profile, the side, top and frontal view, respectively. However, clearly, the the three views are correlated with each other, and constructing a global model comprising all three views allows us to analyse this correlation. In order to do that, we put the 2D shape back to 3D within the respective orthogonal planes, as shown in Figure 3.6. In X-Y-Z frames, the global similarity can be removed by GPA, and the correlation among the three views can be revealed by PCA. Four modes from the global

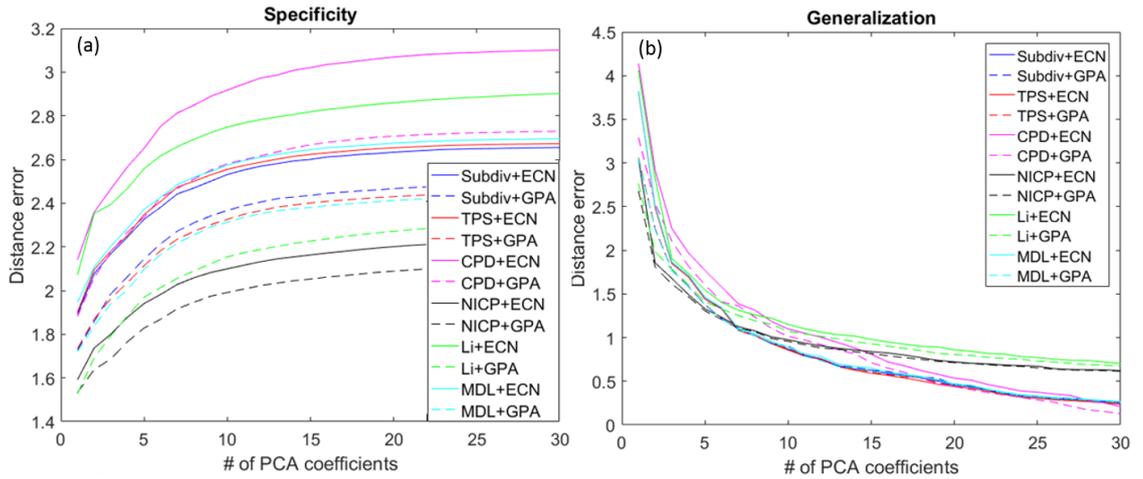


Figure 3.4: (a) specificity; (b) generalization. A lower error is better for both metrics. Alignment methods: ECN, GPA. Correspondence methods: Subdiv, TPS, NICP, Li [92], CPD, MDL. Please view in colour to distinguish the plots.

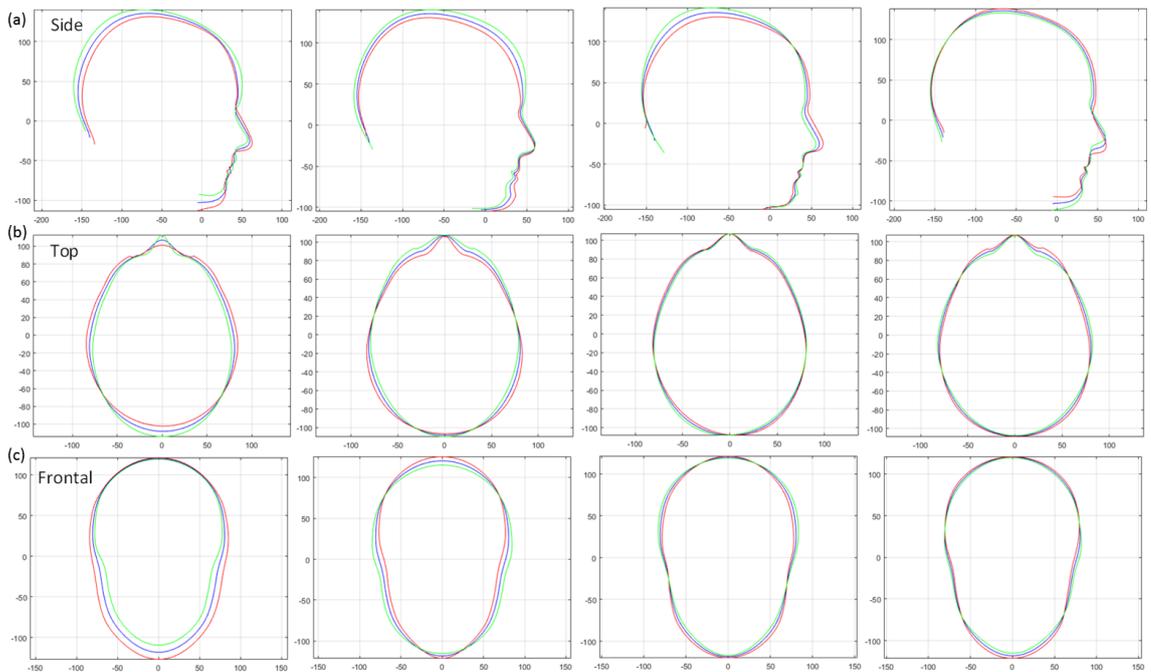


Figure 3.5: Separate models: (a) profile, (b) top and (c) frontal.

model are illustrated in Figure 3.7. Obviously, the variations in Y-Z, X-Z and X-Y within the global model are different from those variations in the separate models in Figure 3.5.

To validate the effectiveness of the global model, we use the parameters from the morphable model to do age and gender classification in the Headspace dataset over 1212 subjects. Using the demographic information (metadata) within the dataset we train a

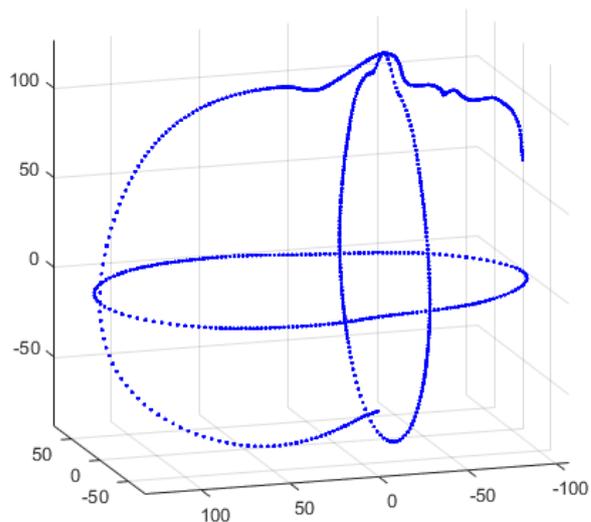


Figure 3.6: Three orthogonal profiles placed within a 3D frame.

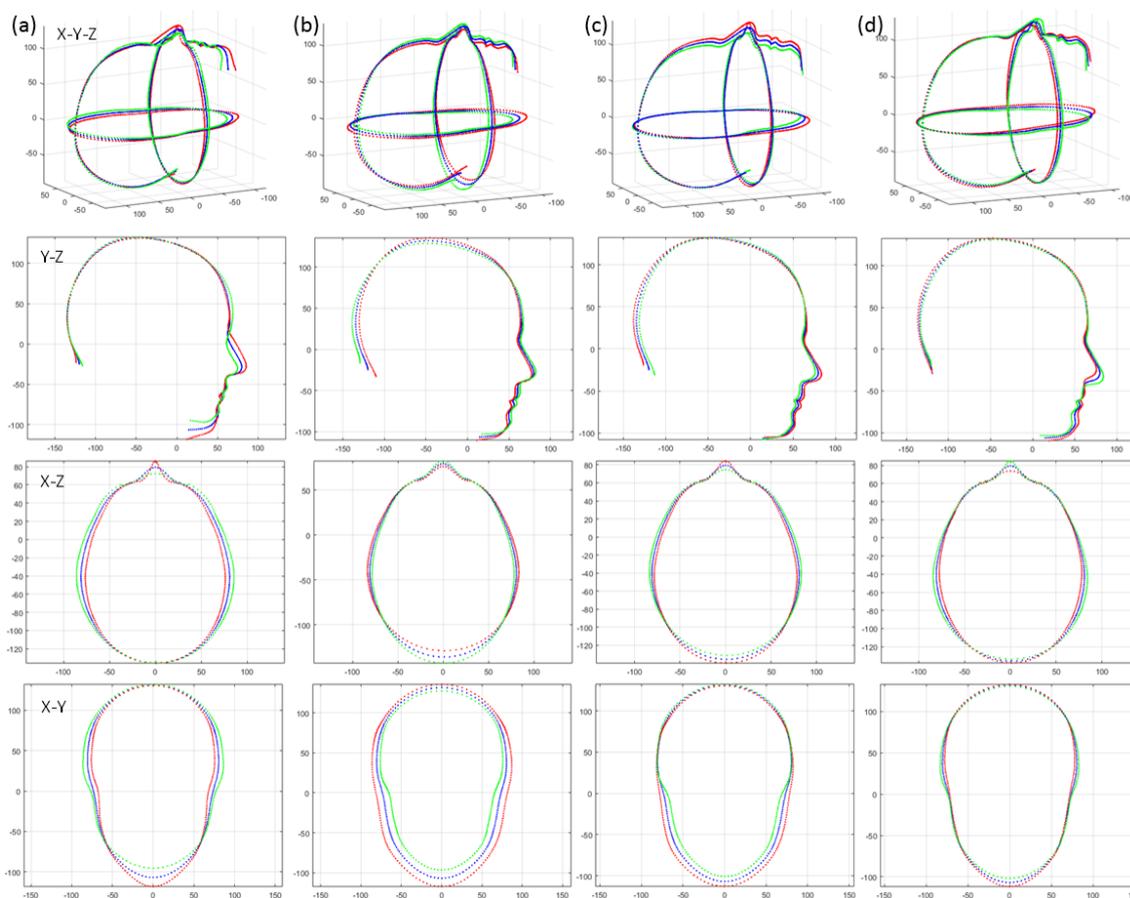


Figure 3.7: Global model, showing the first four modes (a) – (d) of shape variation.

Table 3.1: Gender classification.

<b>Models</b>	<b>Precision</b>	<b>Recall</b>	<b>F-score</b>
Top	0.64	0.65	0.64
Frontal	0.73	0.73	0.73
Profile	0.77	0.77	0.77
<b>Global</b>	<b>0.79</b>	<b>0.79</b>	<b>0.79</b>

Table 3.2: Age classification.

<b>Models</b>	<b>Precision</b>	<b>Recall</b>	<b>F-score</b>
Top	0.72	0.72	0.72
Frontal	0.71	0.71	0.71
Profile	0.73	0.73	0.73
<b>Global</b>	<b>0.75</b>	<b>0.76</b>	<b>0.75</b>

Support Vector Machine (SVM) classifier for each model, which maps the corresponding shape vectors to the gender groups and four age groups (0–11, 12–21, 22–60 and over 60). To measure the classification accuracy, we use the classifier to predict the age bracket and the gender for the test subjects via a 10-fold cross-validation evaluation so that no test subject ever appears in the classifier’s training set. As can be seen in Tables 3.1 and 3.2, the global model has the best performance in both gender and age classification, and the side profile is slightly stronger than the other two separate profiles, as may be expected from the larger curvature variability within that profile.

### 3.3 Summary

We have presented a fully-automatic, general and powerful head profile modelling pipeline that can extract 2D profiles from three orthogonal planes and build both separate and global (combined) morphable models of these profiles. We have thoroughly evaluated two profile alignment methods and six correspondence methods, giving twelve possible

model-building systems enabling the research community to compare and contrast performance across these variants using three different metrics: compactness, specificity and generalisation. Texture based 3D pose normalisation and facial landmarking are applied to extract the sagittal profile from raw 3D scans. Automatic profile annotation, subdivision and registration methods are used to establish dense correspondence among sagittal profiles. ECN builds more compact side profile models when compared to GPA. Subdivision, TPS and MDL with ECN are recommended for a more compact side profile model, while NICP with GPA is recommended to capture more specificity. NICP with GPA is able to generate better reconstructions of unseen profiles when fewer than 7 model dimensions are used. If using between 7 and 20 model dimensions, TPS with ECN is recommended for a better generalisation ability. When more than 20 model dimensions are used, CPD with GPA builds a model with better ability to reconstruct unseen examples. We make our profile models and the data employed to build them publicly available for the research community, enabling others to build models with their own algorithms and compare them to ours.

# Chapter 4

## 3D Statistical Modelling Pipeline

Our 3DMM training pipeline, illustrated in Figure 4.1, operates in three main functional blocks. These are outlined below.

**Data Pre-processing:** We use automatic 2D landmarking and map to 3D using the known 2D-to-3D registration supplied by the 3D camera system. These 3D landmarks can then be used for pose normalisation and template adaptation (personalisation of the template).

**Dense Correspondence:** A collection of 3D scans are reparametrised into a form where each scan has the same number of points joined into a triangulation that is shared across all scans. Furthermore, the semantic or anatomical meaning of each point is shared across the collection, as defined by a template mesh.

**Alignment and Statistical Modelling:** The collection of scans in dense correspondence are subjected to Generalised Procrustes Analysis (GPA) to remove similarity effects (rotation, and translation), leaving only shape information. The processed meshes are statistically analysed, typically with Principal Component Analysis (PCA), generating a 3DMM as a linear basis of shapes. This allows for the generation of novel shape instances.

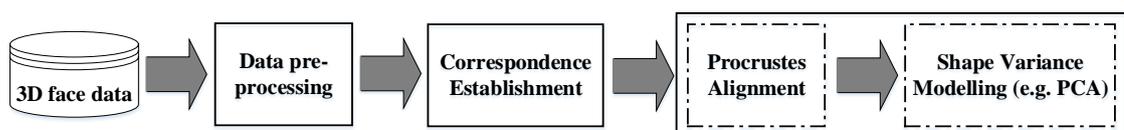


Figure 4.1: The flowchart of 3DMM construction pipeline.

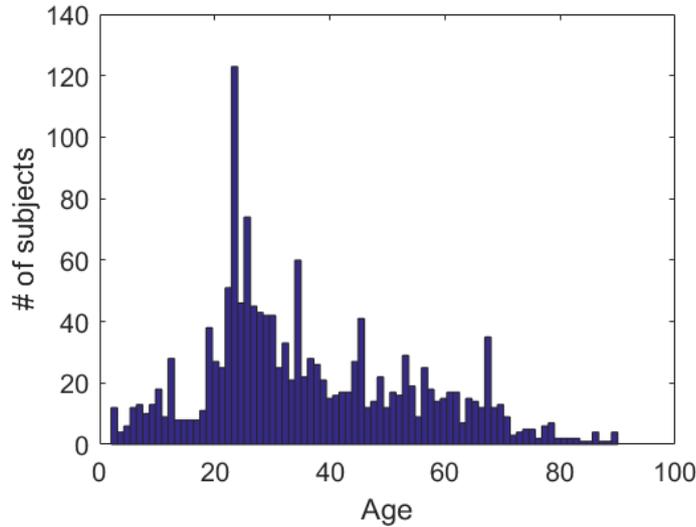


Figure 4.2: Age distribution of subjects.

Each of these functional blocks is detailed in the following four sections, as follows. In section 4.1, we give an overview of the first full head dataset publicly available for academic use. Section 4.2 describes data pre-processing including pose normalisation and 3D facial landmark detection. Section 4.3 presents the correspondence establishment method using an adaptive template. Similarity alignment and statistical modelling are included in section 4.4. Then, section 4.5 presents a texture mapping method that can capture the same resolution as the raw texture image. We compare our 3DMM construction pipeline with other competing construction pipelines in section 4.6 for evaluation of correspondence, and in section 4.7 for evaluation of the resulting 3DMMs. We include applications of our 3DMM in section 4.8 for age regression, cranosynostosis operation outcome assessment, and craniofacial anthropometric measurements. Finally section 4.9 concludes the chapter.

## 4.1 Overview of Headspace Dataset

The Headspace dataset was collected by the Alder Hey Hospital Craniofacial Unit (Liverpool, UK). This unit performs surgery for patients with craniofacial conditions. The aim of such surgery is to protect vital structures (brain, eyes, airway) and normalise appearance, such that patients obtain maximal quality of life through optimal functioning and

avoidance of stigmatisation because of their conditions. Toolsets to define normal appearance and assist with planning of surgery and define optimal outcomes from a range of surgical interventions are currently either crude or lacking. Therefore, the unit collected a large, high quality 3D image dataset of 1519 human heads, with the aim of developing an understanding of the normal variations in the human cranium and face, thus informing software tools for surgical planning and outcome assessment.

The data was collected over a wide age range of 1 to 89 years (see Figure 4.2), but is somewhat unbalanced, with the highest frequency of participants in their 20s. However it is well-balanced in gender. The dataset also includes information like eye color (33.36% brown, 46.38% blue, 19.89% green and 0.37% other) and ethnicity (90% White, 5.3% Asian, 2.7% Mixed Heritage, 1% Black and 1% other). We have made this full head dataset publicly available for research purposes.

## 4.2 Data Pre-processing

Our overall method requires 3D landmark positions, but there are more 2D landmarks available than 3D, and the higher spatial frequency in 2D allows for more accurate localisations. Therefore we use 2D landmarks and project to 3D using the known 2D to 3D registration in the raw data. However, the most accurate 2D landmarking is view-dependent, with some poses being more desirable than others, such as frontal. Since in our dataset, the poses of the 3D scans are not fixed to be frontal, it needs data pre-processing for 3D scan frontalisation. Data preprocessing of the raw 3D scan serves to place the data in a frontal pose, which also allows us to get a complete and accurate set of automatic 3D landmark positions, for every 3D image, that correspond to a set of manually-placed landmarks on the template (this is done only once). This preprocessing comprises five sub-stages, as shown in Figure 4.3: (i) 2D landmarking, (ii) projection to 3D landmarks, (iii) pose normalisation (iv) synthetic frontal 2D image landmarking and (v) projection to 3D landmarks.

Thus there are two stages of landmarking, the first to approximately normalise pose, the second to localise landmarks that correspond to the template mark up, both for fine alignment and template adaptation purposes.

### Data preprocessing

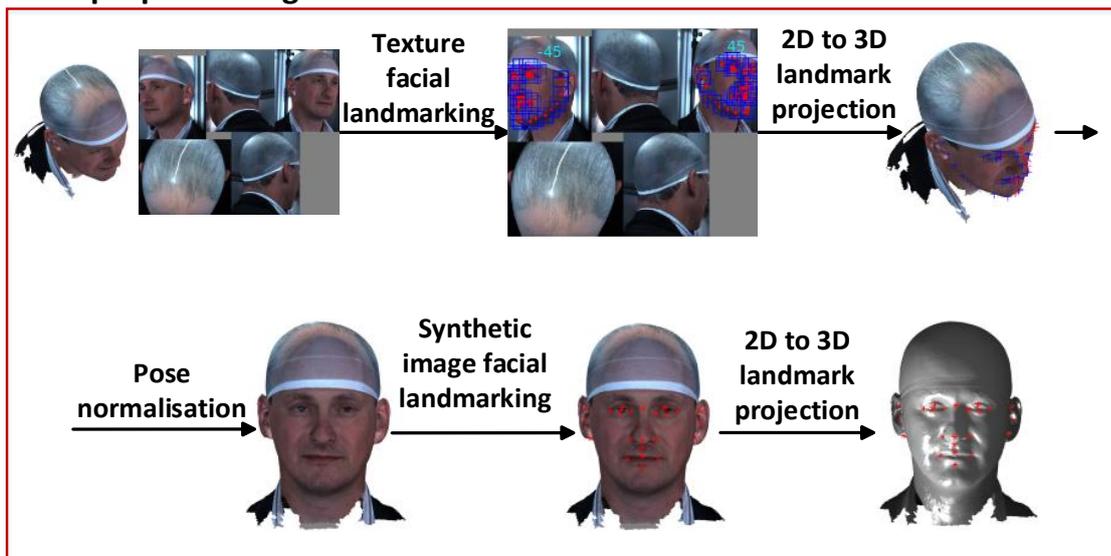


Figure 4.3: Top-left shows raw data from the 3D camera: the textured 3D shape in its original pose, and the composite 5-view 2D colour-texture image. Pre-processing of this raw data comprises five sub-stages: (i) 2D landmarking, (ii) projection to 3D landmarks, (iii) pose normalisation (iv) synthetic frontal 2D image landmarking and (v) projection to 3D landmarks.

We first use the ‘Mixture of Trees’ method of Zhu and Ramanan [157] to localise 2D facial landmarks on the raw 5-view composite texture image. In particular, the mixture we use has 13 landmark tree models for 13 different yaw angles of the head. Two face detections are found, of approximately 15 degrees and 45 degrees yaw from the frontal pose, corresponding to the left and right side of the face respectively. The detected 2D landmarks are then projected to 3D using the OBJ texture coordinates in the raw data.

Given that we know where all of these 3D landmarks should be for a frontal pose, it is possible to do standard 3D pose alignment in a scale-normalised setting [47].

We automatically learn how to orientate each of the trees to frontal pose, based on their 3D structure. To do this, we apply Generalised Procrustes Analysis (GPA) to each collection of 3D trees and find the nearest-to-mean tree shape in a scale-normalised setting. We then apply a 3D face landmarker [43] to the 3D data of the nearest-to-mean tree shape, which generates a set of 14 landmarks with clear semantic meaning. Finally, we find the alignment that moves the symmetry plane of these 14 landmarks to the Y-Z plane

and the nasion directly above the subnasale. To complete the training phase, the mean 3D tree points for each of the 13 trees are then carried into this canonical frontal pose using the same rotation, and are used as reference points for the frontal pose normalisation of the 3D trees.

In around 1% of the dataset, only one tree is detected and that is used for pose normalisation, and in the rest 2-3 images are detected. In the cases where 3 trees are detected, the lowest scoring tree is always false positive and can be discarded. For the remaining two trees, a weighted combination of the two rotations is computed using quaternions, where the weighting is based on the mean Euclidean error to the mean tree, in the appropriate tree component.

After we have rotated the 3D image to canonical frontal view, we wish to generate a set of landmarks that are accurate and correspond to the set marked up on the template. This is the set related to the central tree (0 degrees yaw) in the mixture. After these 2D facial landmarks are extracted, they are again projected onto 3D mesh.

The work of Zhou et al. [155] shows that an AAM with SIFT features has excellent performance in ear landmark detection. We just need three non-colinear ear landmarks for ear alignment, which is a simpler task than that in [155]. For each subject, we generate two synthetic images in profile view by rotating 90 degrees and -90 degrees yaw. We then use an AAM with SIFT features to detect the three ear landmarks on each synthetic image. We then project the 2D landmarks onto 3D mesh. The detected facial landmarks and ear landmarks are shown in Figure 4.4.

## 4.3 Correspondence Establishment

We present a new pipeline in fully-automatic non-rigid 3D shape registration by integrating several powerful ideas from the computer vision and graphics. These include Iterative Closest Points (ICP) [18], Coherent Point Drift (CPD) [100], and mesh editing using the Laplace-Beltrami (LB) operator [120]. We also provide comparisons of the latter approach with the use of Gaussian Processes (GPs) [67].

Our contributions include: 1) an adaptive shape template method to accelerate the convergence of registration algorithms and achieve a better final shape correspondence;

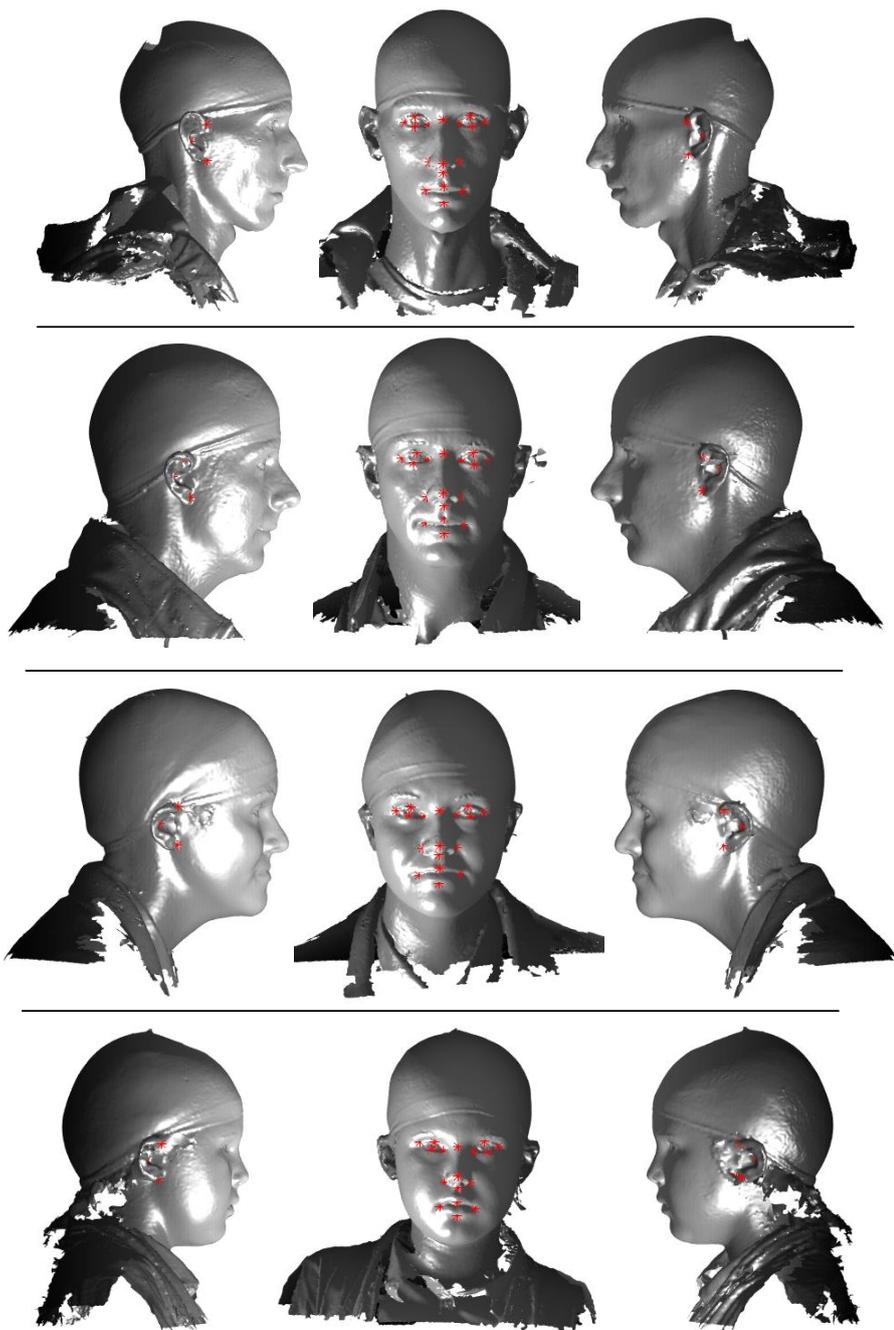


Figure 4.4: Landmarks detection results: 1st column - right ear landmarks; 2nd column - facial landmarks; 3rd column - left ear landmarks.

and 2) a new iterative registration method that combines ICP with CPD to achieve a more stable and accurate correspondence establishment than standard CPD. We call this approach *Iterative Coherent Point Drift (ICPD)*. These two processing stages are illustrated in Figure 4.5, and are presented in Section 4.3.1 and Section 4.3.2 respectively.

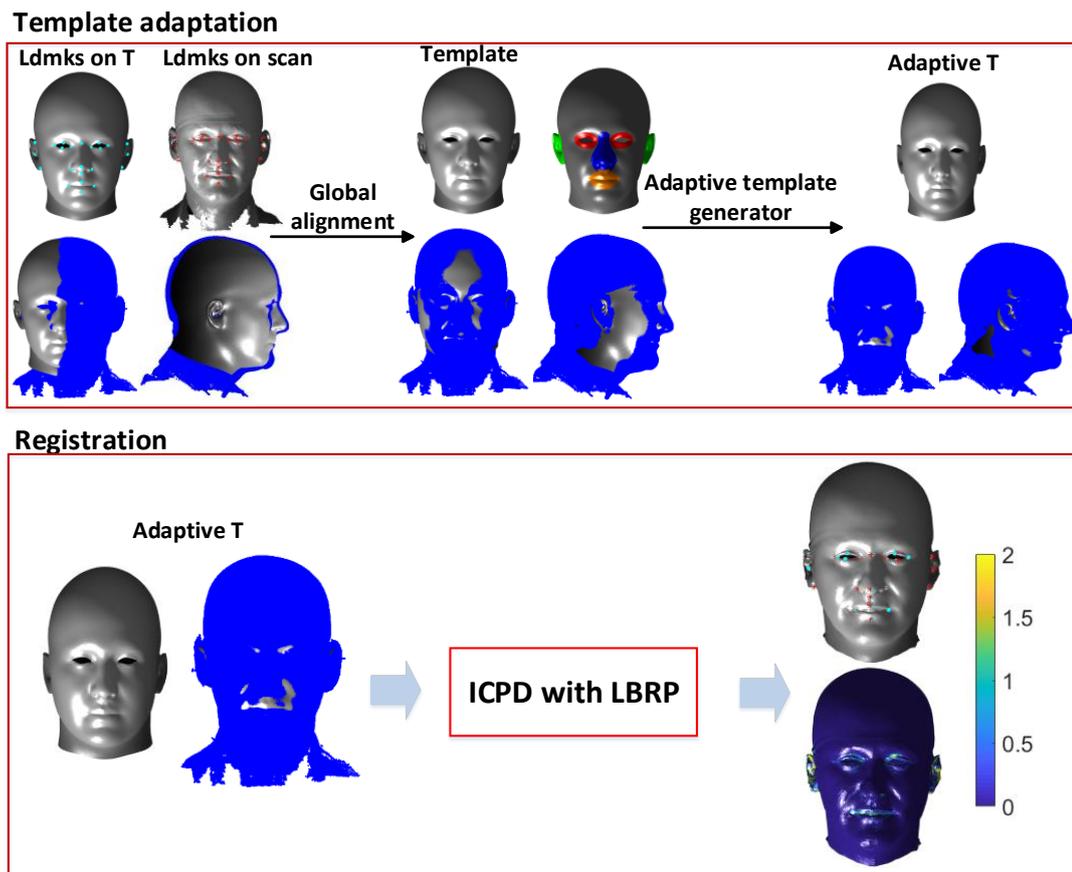


Figure 4.5: Template morphing framework: the upper row - template adaption process; the lower row - registration for dense correspondence using ICPD with Laplace-Beltrami regularised projection.

Figure 4.6 is a qualitative illustration of a typical result where our method achieves a more accurate correspondence than standard CPD. Note that the landmarks in our method are almost exactly the same position as their corresponding ground-truth points on the raw 3D scan. Even though standard CPD-affine is aided by Laplace-Beltrami regularised projection (LBRP, a component of our new pipeline), the result shows a *squeezed* face around the eye and mouth regions and the landmarks are far away from their corresponding ground-truth positions.

### 4.3.1 Template Adaptation

As shown in Figure 4.5, template adaptation consists of two sub-stages: (i) global alignment followed by (ii) dynamically adapting the template shape to the data. For global

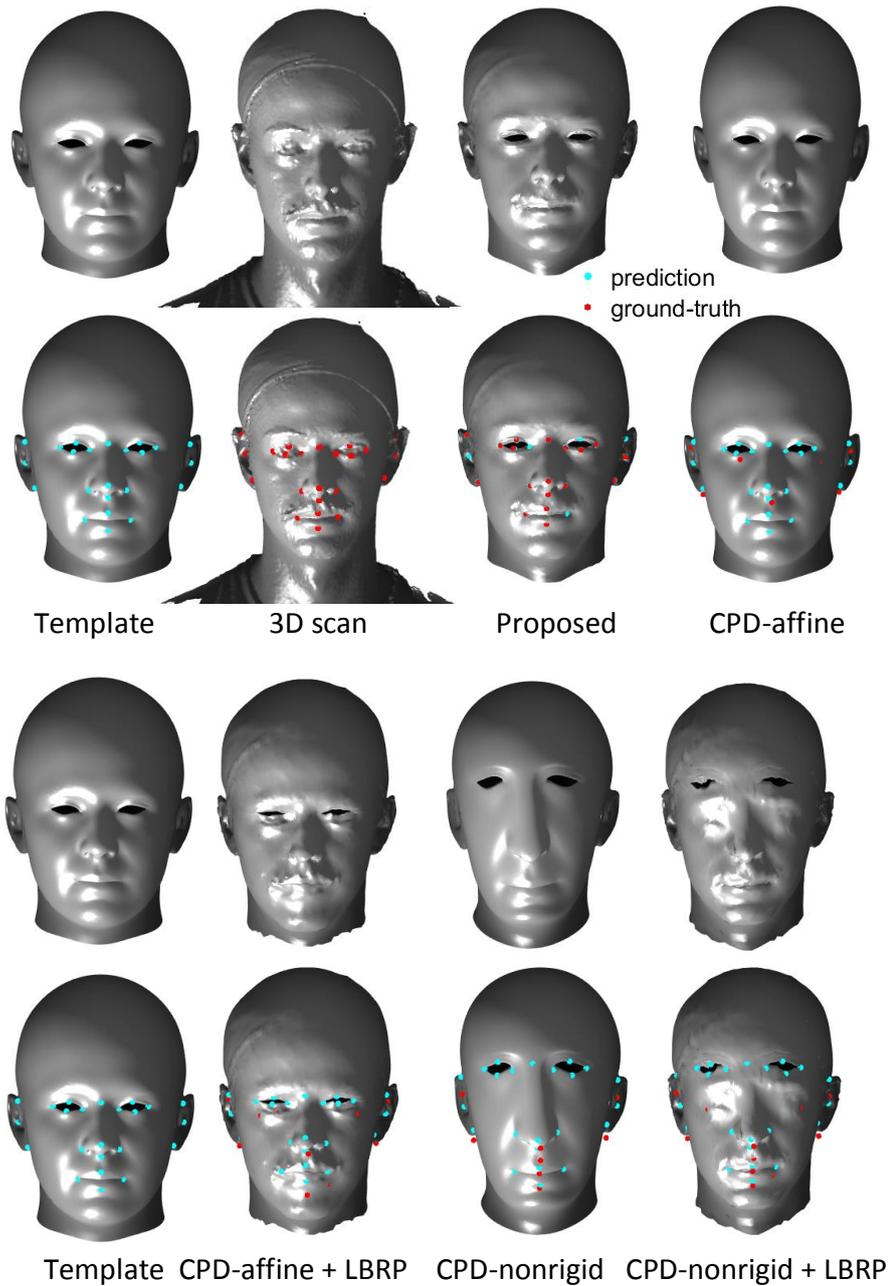


Figure 4.6: Proposed method compared with standard CPD. Ground truth points on target raw 3D data shown in red, corresponding template points shown in cyan.

alignment, we manually select the same landmarks on the template as we automatically extract on the raw data (i.e. using the zero yaw angle tree component from [157]). Note that this needs to be done *once only* for some object class and so does not impact on the autonomy of the *online* operation of the framework. Then we align rigidly (without scaling) from the 3D landmarks on raw 3D data to the same landmarks on the template. The

rigid transformation matrix is used for the raw data alignment to the template.

The template is then *adapted* to better align with the raw scan. A better template helps the later registration converge faster and gives more accurate correspondence at the beginning and end of registration. A good template has the same size and position of local facial parts (e.g. eyes, nose, mouth and ears) as the raw scan. This cannot be achieved by mesh alignment alone. We propose two methods to give a better template that is adapted to the raw 3D scan: (1) Laplace-Beltrami mesh editing; (2) Template estimation via posterior GPMMs. For both methods, three ingredients are needed: landmarks on 3D raw data, the corresponding landmarks on template, and the original template.

#### **Laplace-Beltrami Mesh Manipulation:**

We decompose the template into several facial parts: eyes, nose, mouth, left ear and right ear. We rigidly align landmarks on each part separately to their corresponding landmarks on 3D raw data. These rigid transformation matrices are used for aligning the decomposed parts to 3D raw data. The rigidly transformed facial parts tell the original template where it should be. We treat this as a mesh manipulation problem. We use Laplace-Beltrami mesh editing to manipulate the original template towards the rigidly transformed facial parts, as follows: (1) the *facial parts* (fp) of the original template are manipulated towards their target positions - these are rigidly transformed facial parts; (2) all other parts of the original template are moved *as rigidly as possible* [121]. As shown in Figure 4.7, the facial parts include the eyes region, the nose region, the mouth region and the ears region. We manually select these facial parts on the template, which is also a one-shot offline process.

Given the vertices of a template stored in the matrix  $\mathbf{X}_T \in \mathbb{R}^{p \times 3}$  and a better template obtained whose vertices are stored in the matrix  $\mathbf{X}_{bT} \in \mathbb{R}^{p \times 3}$ , we define the selection matrices  $\mathbf{S}_{fp} \in [0, 1]^{l \times p}$  as those that select the  $l$  vertices (facial parts in  $\mathbf{X}_T$  and  $\mathbf{X}_{bT}$ ) from the raw template and a better template respectively. This linear system can be written as:

$$\begin{pmatrix} \lambda \mathbf{L} \\ \mathbf{S}_{fp} \end{pmatrix} \mathbf{X}_{bT} = \begin{pmatrix} \lambda \mathbf{L} \mathbf{X}_T \\ \mathbf{X}_{fp} \end{pmatrix} \quad (4.1)$$

where  $\mathbf{L} \in \mathbb{R}^{p \times p}$  is the cotangent Laplacian approximation to the LB operator [121] and  $\mathbf{X}_{bT}$  is the better template that we wish to solve for. The parameter  $\lambda$  weights the relative influence of the position and regularisation constraints, effectively determining



Figure 4.7: Annotation of facial parts on the template: eyes region - red, nose region - blue, mouth region - orange, and ears region - green.

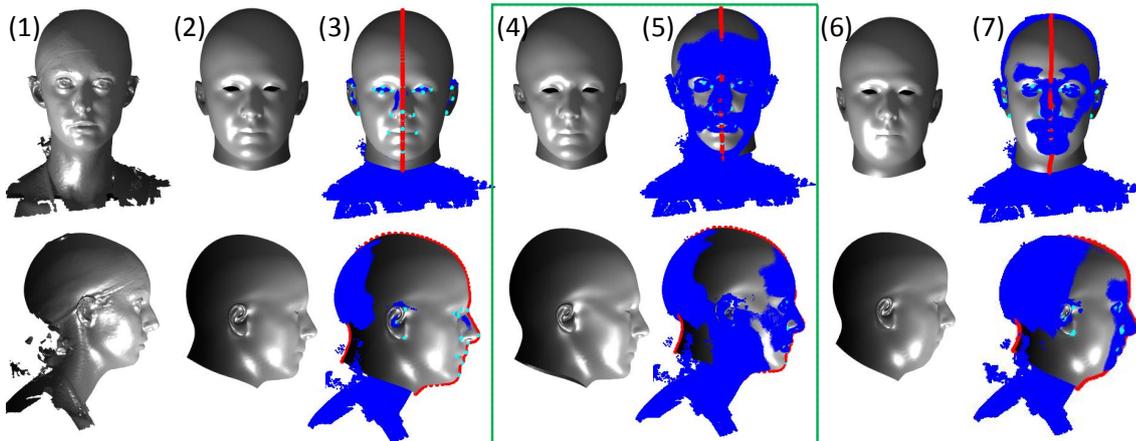


Figure 4.8: (1) raw scan; (2) template with global rigid alignment; (3) 2 compared with the raw scan; (4) adaptive template via LB mesh editing; (5) 4 compared with the raw scan; (6) the mean template estimation via posterior models; (7) 6 compared with the raw scan in (7).

the ‘stiffness’ of the mesh manipulation. As  $\lambda \rightarrow 0$ , the facial parts of the original template are manipulated exactly to the rigidly transformed facial parts. As  $\lambda \rightarrow \infty$ , the adaptive template will only be at the same position as the original template  $X_T$ .

### Template Estimation via Posterior Models:

A common task in shape modelling is to infer the full shape from a set of measurements of the shape. This task can be formalised as a regression problem. The posterior models of Gaussian Process Morphable Models (GPMMs) are regression models of the deformation field. Given partial observations, posterior models are able to answer what is the potential full shape. Posterior models show the points' potential movements when the landmarks are fixed to their target position.

In a GPMM, let  $\{x_1, \dots, x_l\} \in \mathbb{R}^{l \times 3}$  be a fixed set of input 3D points and assume that there is a regression function  $f_0 \rightarrow \mathbb{R}^{p \times 3}$ , which generates a new vector field  $\mathbf{y}_i \in \mathbb{R}^{p \times 3}$  according to

$$\mathbf{y}_i = f_0(\mathbf{x}_i) + \epsilon_i, (i = 1, \dots, n). \quad (4.2)$$

where  $\epsilon_i$  is independent Gaussian noise, i.e.  $\epsilon_i \sim N(0, \delta^2)$ . The regression problem is to infer the function  $f_0$  at the input points  $\{x_1, \dots, x_l\}$ . The possible deformation field  $\mathbf{y}_i$  is modelled using a Gaussian process model  $GP(\mu, k)$  that models the shape variations of a given shape family.

In our case, the reference shape is the original template, the landmarks on the original template are the fixed set of input 3D points. The same landmarks on 3D raw data are the target position of the fixed set of input 3D points. We construct a GPMM  $GP(\mu, k)$  by pre-defining Gaussian kernels to model the shape variations of a shape family. The pre-defined Gaussian kernels follows the same strategy in [67]. The adaptive template is

$$\mathbf{X}_{bT}^i = \mathbf{X}_T + \mathbf{y}_i, (i = 1, \dots, n). \quad (4.3)$$

The mean of  $\mathbf{X}_{bT}^i$  is shown in Figure 4.8 (6) and (7).

### 4.3.2 Iterative Coherent Point Drift

After template alignment and adaptation, we employ a new shape morphing (non-rigid 3D registration) algorithm. The task is to deform and align the template to the target raw 3D scan. Non-rigid Coherent Point Drift (CPD) [100] has better deformation results when partial correspondences are given and we have found that it is more stable and converges better when the template and the raw data have approximately the same number of points. However, the correspondence is often not known before registration. Thus, following

an Iterative Closest Points (ICP) scheme [18], we supply CPD registration with coarse correspondences using ‘closest points’. We refine such correspondences throughout iterations of our *Iterative Coherent Point Drift* (ICPD) approach. ICP assumes that we do not know the correspondence in advance, but we can obtain a better one in an iterative way when starting from a coarse correspondence. There are many methods to compute the correspondence, but we adopt the standard nearest point search approach.

We use the original code package of CPD available online as library calls for ICPD. Other option parameters can be found in the CPD author’s release code. The global affine transformation is used as a small adjustment of correspondence computation. A better correspondence is used as the priors for CPD non-rigid registration.

**Algorithm:** We update the correspondence priors for ICPD iteratively. The correspondence priors are refined in a coarse-to-fine manner following the idea ICP approach. So the convergence of outer loop is the stabilization of correspondence priors. The matlab pseudocode of ICPD is given as:

---

**Algorithm 1** : Nonrigid registration using adaptive template

---

- 1: Solve for the similarity (scaled rigid) transformation between the landmarks on the template and on the scan (data):  
 $[\mathbf{R}, \mathbf{t}, s] \leftarrow \operatorname{argmin} \| (s\mathbf{S}_{ldmkT}\mathbf{X}_T\mathbf{R} + \mathbf{t}) - \mathbf{S}_{ldmkS}\mathbf{X}_{scan} \|$
  - 2: Global alignment of the template using the similarity transformation:  
 $\mathbf{X}'_T = s\mathbf{X}_T\mathbf{R} + \mathbf{t}$
  - 3: Obtain adaptive template  $\mathbf{X}_{bT}$  by solving Equation 4.1
  - 4: Set **flag** = 0,  $\mathbf{X}_{deformed} = \mathbf{X}_{bT}$
  - 5: **while flag** = 0 **do**
  - 6:   Compute the selection matrix  $\mathbf{S}_{old}$  that selects the nearest neighbours of  $\mathbf{X}_{deformed}$  in  $\mathbf{X}_{scan}$
  - 7:   Solve for the optimal affine transformation:  
 $[\mathbf{R}_a, \mathbf{t}_a] \leftarrow \operatorname{argmin} \| (\mathbf{X}_{deformed}\mathbf{R}_a + \mathbf{t}_a) - \mathbf{S}_{old}\mathbf{X}_{scan} \|$
  - 8:   Update the mesh deformation:  $\mathbf{X}_{deformed} \leftarrow \mathbf{X}_{deformed}\mathbf{R}_a + \mathbf{t}_a$
  - 9:   Compute the new selection matrix  $\mathbf{S}_{new}$  that selects the nearest neighbours of  $\mathbf{X}_{deformed}$  in  $\mathbf{X}_{scan}$
  - 10:   Use the *CPD-nonrigid* algorithm to solve for the non-rigid deformation between  $\mathbf{X}_{deformed}$  and  $\mathbf{S}_{new}\mathbf{X}_{scan}$
  - 11:   Update  $\mathbf{X}_{deformed}$  using this non-rigid deformation
  - 12:   Update the selection matrix  $\mathbf{S}_{new}$
  - 13:   **if**  $f_{diff}(\mathbf{S}_{old}, \mathbf{S}_{new}) < 0.001 * n$ , where  $n$  is the number of points in template,  $f_{diff}$  indicates the number of different entries when comparing  $\mathbf{S}_{old}$  and  $\mathbf{S}_{new}$ , **then**
  - 14:     **flag** = 1
  - 15:   **end if**
  - 16: **end while**
- 

where  $\mathbf{S}_{ldmkS}$  and  $\mathbf{S}_{ldmkT}$  are the selection matrices of the landmarks on 3D scan  $\mathbf{X}_{scan}$  and morphing template  $\mathbf{X}_T$ .  $[\mathbf{R}, \mathbf{t}, s]$  are the rotation, translation and scaling, respectively.

$[\mathbf{R}_a, \mathbf{t}_a]$  define the affine transformation.

The rigid, affine and nonrigid deformation algorithms are from the original code package of CPD.  $f_{diff}$  indicates the number of different entries when comparing  $\mathbf{S}_{old}$  and  $\mathbf{S}_{new}$ . The qualitative output of ICPD is very smooth, a feature inherited from standard CPD. A subsequent regularised point projection process is required to capture the target shape detail, and this is described next.

### 4.3.3 Laplace-Beltrami Regularised Projection

When ICPD has deformed the template close to the scan, point projection is required to eliminate any (normal) shape distance error. Point projection is a potentially fragile process. If the scan data is incomplete or noisy then projecting vertices from the deformed template to their nearest vertex or surface position on the scan may cause large artefacts. Again, we overcome this by treating the projection operation as a mesh editing problem with two ingredients. First, position constraints are provided by those vertices with mutual nearest neighbours between the deformed template and raw scan. Using mutual nearest neighbours reduces sensitivity to missing data. Second, regularisation constraints are provided by the LB operator which acts to retain the local structure of the mesh. We call this process *Laplace-Beltrami regularised projection* (LBRP), as shown in the registration framework in Figure 4.5.

We write the point projection problem as a linear system of equations. Given the vertices of a scan stored in the matrix  $\mathbf{X}_{scan} \in \mathbb{R}^{n \times 3}$  and the deformed template obtained by CPD whose vertices are stored in the matrix  $\mathbf{X}_{deformed} \in \mathbb{R}^{p \times 3}$ , we define the selection matrices  $\mathbf{S}_1 \in [0, 1]^{m \times p}$  and  $\mathbf{S}_2 \in [0, 1]^{m \times n}$  as those that select the  $m$  vertices with nearest neighbours from deformed template and scan respectively. This linear system can be written as:

$$\begin{pmatrix} \lambda \mathbf{L} \\ \mathbf{S}_1 \end{pmatrix} \mathbf{X}_{proj} = \begin{pmatrix} \lambda \mathbf{L} \mathbf{X}_{deformed} \\ \mathbf{S}_2 \mathbf{X}_{scan} \end{pmatrix} \quad (4.4)$$

where  $\mathbf{L} \in \mathbb{R}^{p \times p}$  is the cotangent Laplacian approximation [121] to the LB operator and  $\mathbf{X}_{proj} \in \mathbb{R}^{p \times 3}$  are the projected vertex positions that we wish to solve for. The parameter  $\lambda$  weights the relative influence of the position and regularisation constraints, effectively determining the ‘stiffness’ of the projection. As  $\lambda \rightarrow 0$ , the projection tends towards

nearest neighbour projection. As  $\lambda \rightarrow \infty$ , the deformed template will only be allowed to rigidly transform.

## 4.4 Similarity Alignment and Statistical Modelling

We use GPA to align our deformed templates before applying PCA-based statistical modelling. This generates a 3DMM as a linear basis of shapes, allowing for the generation of novel shape instances. Note that we may select all of the full head template vertices for this modelling, or any subset. For example, we select the cranial vertices when we build models to analyse cranial medical conditions.

In many applications, vertex resolution is not uniform across the mesh. For example, we may use more vertices to express detail around facial features of high curvature. However, standard PCA attributes the same weight to all points in its covariance analysis, making it difficult to handle shape variance modelling over such meshes. To counter this, we employ weighted PCA in our statistical modelling.

### 4.4.1 Weighted PCA

The first step of Weighted PCA (WPCA) is to construct the symmetric positive-definite matrix  $\mathbf{Y}^T \mathbf{M} \mathbf{Y}$  for  $\mathbf{Y} = \mathbf{X} - \bar{\mathbf{X}}$ , where the mass matrix  $\mathbf{M}$  is a sparse symmetric mass matrix based on the areas of the mesh triangles [141]. The first  $k$  eigenvectors  $\mathbf{V}_i^k$  solving

$$\mathbf{Y}^T \mathbf{M} \mathbf{Y} \mathbf{V}_i^k = \lambda_i \mathbf{V}_i^k \quad (4.5)$$

are computed. From [77], the relation between the left singular vector  $\mathbf{U}_i^k$  and the right singular vector  $\mathbf{V}_i^k$  of SVD decomposition is  $\mathbf{U}_i^k = (\frac{1}{\sqrt{\lambda_i^k}}) \mathbf{M}^{\frac{1}{2}} \mathbf{Y} \mathbf{V}_i^k$ . For  $1 \leq i \leq k$ , we infer that the weighted PCA  $\mathbf{\Psi}_i^k$  is obtained by multiplying the eigenvectors with  $\mathbf{Y}$  and normalisation(w.r.t $\|\cdot\|$ ) of the result:

$$\mathbf{\Psi}_i^k = \mathbf{M}^{-\frac{1}{2}} \mathbf{U}_i^k = \left( \frac{1}{\sqrt{\lambda_i^k}} \right) \mathbf{Y} \mathbf{V}_i^k \quad (4.6)$$

### 4.4.2 Reconstruction

In an analogous fashion to standard PCA, WPCA can also be used for reconstructing shapes.

**Complete data reconstruction:** Given the average  $\bar{\mathbf{X}}$ , the basis matrix  $\mathbf{U}_i^k$ , the mass matrix  $\mathbf{M}$  and some (possibly unseen) shape  $\mathbf{X} \in \mathbb{R}^{3n}$ , we first compute the zero mean  $\mathbf{U} = \mathbf{X} - \bar{\mathbf{X}}$ . Then  $\mathbf{U}$  is projected to WPCA space:

$$\mathbf{P}^k : \mathbf{U} \rightarrow \mathbf{P}^k \boldsymbol{\Psi}^k = \mathbf{P}^k \mathbf{M}^{-\frac{1}{2}} \mathbf{U}_i^k \quad (4.7)$$

Finally we are able to reconstruct  $\mathbf{X}$  by:

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{P}^k \boldsymbol{\Psi}^k = \bar{\mathbf{X}} + \mathbf{P}^k \mathbf{M}^{-\frac{1}{2}} \mathbf{U}_i^k \quad (4.8)$$

**Incomplete data reconstruction:** Given partial data, we can divide the shape components into two parts: one for the partial data  $\boldsymbol{\Psi}_a$  and the other for missing data  $\boldsymbol{\Psi}_b$  [3]. The zero mean for the partial data is  $\mathbf{U}_a = \mathbf{X}_a - \bar{\mathbf{X}}_a$ :

$$\mathbf{P} : \mathbf{U}_a \rightarrow \mathbf{P} \boldsymbol{\Psi}_a = \mathbf{P}_a \mathbf{M}^{-\frac{1}{2}} \mathbf{U}_a \quad (4.9)$$

Then  $\mathbf{P}$  can be used to reconstruct the complete data. The aim is to model the remaining flexibility of the variable points  $\mathbf{X}_a$  when the fixed points  $\mathbf{X}_b$ . The deformation is give as  $\mathbf{P} \boldsymbol{\Psi}_a$  and  $\mathbf{P} \boldsymbol{\Psi}_b$ . We formulate the aim as a constrained maximization problem:

$$\max \mathbf{P} \boldsymbol{\Psi}_a \boldsymbol{\Psi}_a^T \mathbf{P}^T, \text{ subject to } \mathbf{P} \boldsymbol{\Psi}_b \boldsymbol{\Psi}_b^T \mathbf{P}^T = c \quad (4.10)$$

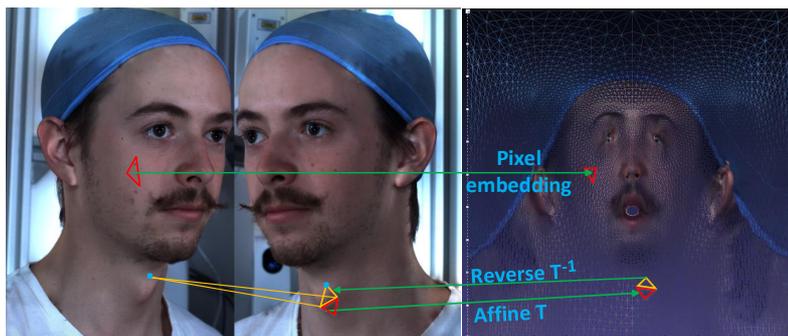
where  $c$  quantifies the amount of change allowed in the fixed shape principle components. Introducing a Lagrangian multiplier  $\mu$  and differentiating with respect to  $\mathbf{P}$  leads to the eigenvalue decomposition of:

$$\mathbf{P} \boldsymbol{\Psi}_a \boldsymbol{\Psi}_a^T = \mu \mathbf{P} \boldsymbol{\Psi}_b \boldsymbol{\Psi}_b^T \quad (4.11)$$

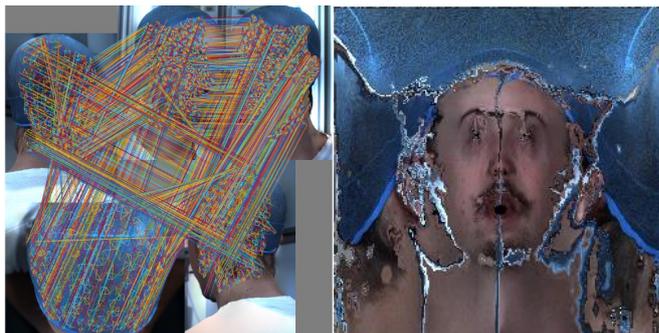
where for each parameter in  $\mathbf{P}$ ,  $\|\mathbf{P}_i \boldsymbol{\Psi}_b\|^2 = \mathbf{P}_i \boldsymbol{\Psi}_b \boldsymbol{\Psi}_b^T \mathbf{P}_i^T$ .

## 4.5 Texture Map

It is preferable to store texture information in a UV space texture map, where resolution is unconstrained, rather than store only per-vertex colours, where resolution is limited by mesh resolution. To do so requires the texture information from each scan to be transformed into a standard UV texture space for which the embedding of the template is



(1) Pixel embedding and affine transformation refinement



(2) Before



(3) After

Figure 4.9: Texture mapping flow: (1) Pixel embedding, (2) UV coordinates of mesh facets before affine refinement, here we are unable to use facets that straddle viewpoints, (3) After affine refinement, all facet UV coordinates can be used.

known. The key to obtaining a high quality texture map is embedding all the pixels in one face from the texture image to its corresponding face in the template UV faces (see Figure 4.9 (1)). Compared to a per-vertex color-texture map, a pixel embedding texture map keeps all the pixels in each template mesh face, thus capturing more texture detail.

After template morphing, the deformed template has the same number of points joined into a triangulation that is shared across all scans. Thus in UV coordinates, UV faces of

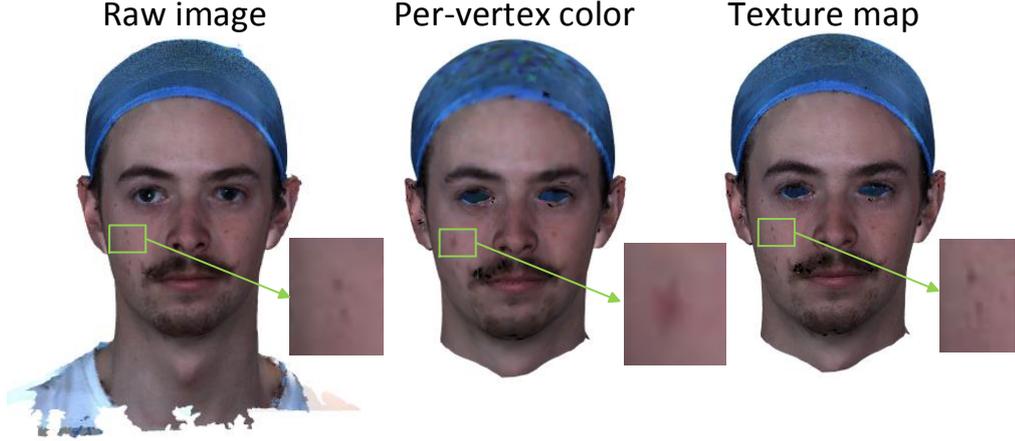


Figure 4.10: Per-vertex color texture (centre) vs. texture map (right).

the morphed template are shared with the template. Given the morphed vertex positions  $\mathbf{X}_{\text{morphed}} \in \mathbb{R}^{p \times 3}$  from template morphing stage, we can first compute the UV coordinates for each point of the morphed template in original texture image:

$$[\mathbf{u}, \mathbf{v}] = g(\mathbf{S}_3 \mathbf{X}_{\text{scan}}) \quad (4.12)$$

where  $\mathbf{S}_3 \in \mathbb{R}^{p \times n}$  is the selection matrix that select the  $p$  vertices with nearest neighbours from morphed template  $\mathbf{X}_{\text{morphed}}$  to the scan  $\mathbf{X}_{\text{scan}}$  and  $g$  is the UV coordinates mapping from the raw mesh to texture image. The UV coordinates mapping from texture image to raw mesh is a surjection but not an injection. Thus the points from the raw mesh may have several sets of UV coordinates  $[\mathbf{u}, \mathbf{v}]_{\text{cand}}$  in the texture image, depending on the number of capture viewpoints. To overcome this, we minimise the face area of UV face  $\mathbf{UV}_{\text{faces}}^k$  to find the exact UV coordinates:

$$[\mathbf{u}_i^*, \mathbf{v}_i^*] = \arg \min_{\mathbf{u}_i, \mathbf{v}_i} h([\mathbf{u}_i, \mathbf{v}_i], [\mathbf{u}_{j1}, \mathbf{v}_{j1}], [\mathbf{u}_{j2}, \mathbf{v}_{j2}]) \quad (4.13)$$

where  $[\mathbf{u}_i, \mathbf{v}_i] \in [\mathbf{u}, \mathbf{v}]_{\text{cand}}$ ,  $[\mathbf{u}_{j1}, \mathbf{v}_{j1}]$  and  $[\mathbf{u}_{j2}, \mathbf{v}_{j2}] \in \mathbf{UV}_{\text{face}}^k$ , and  $h$  is the face area computation.

Due to multiple capture viewpoints, the minimization of face area is not enough to guarantee that all the UV coordinates are in the image of same view (see Figure 4.9 (2)). To overcome this, we employ affine transformations to refine the UV coordinates. If the UV coordinates in one mesh face are placed in different views, we compute the affine

transformation  $\mathbf{T}$  from its adjacent face to the corresponding face in the template UV faces. Then its corresponding face is inverted by  $\mathbf{T}^{-1}$  to find the correct position (see blue point in Figure 4.9 (1)). The outcome of affine transformation refinement is shown in Figure 4.9 (2), (3). As shown in Figure 4.10, the quality of texture map improves a lot such that the freckles can be seen from the texture rendering.

## 4.6 Evaluation of Correspondences

We evaluated the proposed template morphing algorithms using both the BU3D dataset [149] and the Headspace dataset [1]. Section 4.6.1 presents the validation of the template adaptation approach. In this thesis, we proposed three registration methods: (1) the method detailed in this chapter, (2) the symmetric deformation algorithm detailed in chapter 4, and (3) our earlier method described in Appendix B and [47]. These are compared in section 4.6.2. In section 4.6.3, we present the qualitative and quantitative evaluation of the correspondence establishment with other competing template morphing approaches.

### 4.6.1 Validation of Template Adaptation

In order to validate the effectiveness of each key step in the proposed registration pipeline, we first remove the process of template adaption from the pipeline and evaluate performance. We then replace this and remove LB regularised projection from the proposed framework and again evaluate performance qualitatively and quantitatively, comparing both modified pipeline cases with the full pipeline. Typical results for a child in the Headspace dataset are shown in Figure 4.11. After pure rigid alignment without template adaptation, the nose of the template is still bigger than the target. As can be seen in Figure 4.11 (3), the nose ends up with a bad deformation result. The same problem happened in the ear. Without LB regularised projection, shown in Figure 4.11 (4), it fails in capturing the shape detail compared with the proposed full pipeline. The adaptive template improves the correspondence accuracy in local regions, while the LB regularised point projection helps in decreasing the correspondence error in the normal direction.

Using the BU3D dataset for quantitative validation, we compared the performance of (i) the proposed ICPD registration, (ii) ICPD with an adaptive template using LB mesh

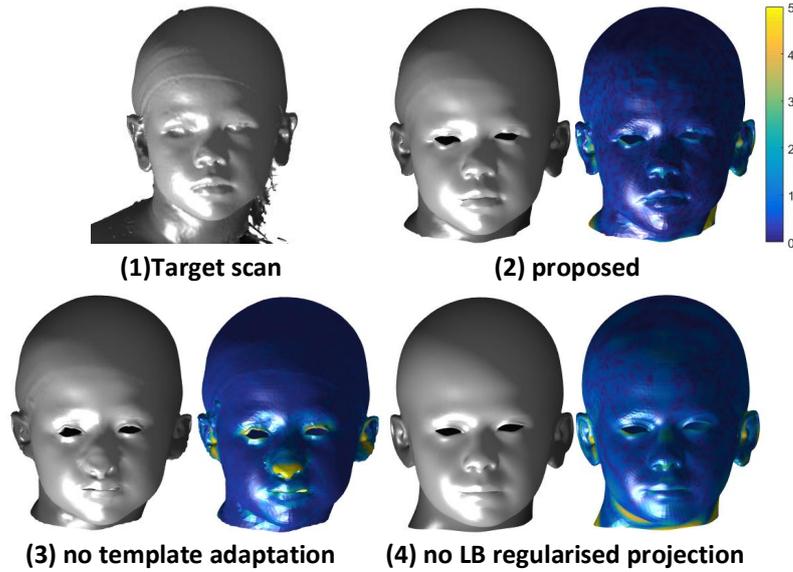
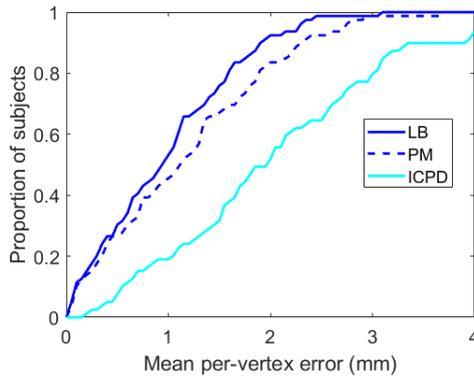


Figure 4.11: Internal comparison of approaches: (1) target scan; (2) proposed full pipeline method; (3) template adaptation removed; (4) LB regularised projection removed. Error map shown in mm.



**Table.** Convergence of ICPD

	ICPD	LB	PM
ICPD loops	6.47	3.52	3.74
Time (s)	831.35	426.13	434.53

Figure 4.12: Improvement in correspondence and convergence performance when using adaptive templates: 1) ICPD without an adaptive template (cyan); 2) ICPD with LB-based adaptive template (blue); 3) ICPD with adaptive PM-based template (blue dashed).

manipulation and (iii) ICPD with an adaptive template, using a posterior model (PM). The mean per-vertex error is computed between the registration results and their ground-truth. The number of ICPD iterations and computation time is recorded, when using the same computation platform. The per-vertex error plot in Figure 4.12 illustrates that the adaptive template improves the correspondence accuracy of ICPD. The number of ICPD iterations

and computation time is significantly decreased by the adaptive template method. In particular adaptive template using LB mesh manipulation has better performance than adaptive template using a posterior model. Thus, we employ an adaptive template approach using LB mesh manipulation for later experiments.

### 4.6.2 Comparison Between Proposed Approaches

During this thesis work, we proposed three different template morphing algorithms, as discussed earlier. We believe that the template morphing framework in this chapter is our current best system. To validate this, we compare our proposed template morphing algorithms: ICPD with adaptive template [46] presented in this chapter, *hierarchical parts-based* CPD-LB [47] (also in Appendix B) and symmetry-aware CPD [48] (also in Chapter 4). Figure 4.13 illustrates that the eye region and mouth region can have a small over-fitting problem in morphing when using either *hierarchical parts-based* CPD-LB or symmetry-aware CPD. The third row in Figure 4.13 shows that ICPD with adaptive template gives a better morphing in ear region where outliers exists in the raw data.

To evaluate the three approaches quantitatively, we use 14 manual facial landmarks over 100 subjects to measure landmark error of each template deformation method. The mean per-vertex error is calculated by the mean nearest point error between the morphed template and its corresponding 3D mesh. As can be seen from Figure 4.14 (1), ICPD with adaptive template is much better than the other two in terms of mean per-vertex error. Figure 4.14 (2) shows that ICPD with adaptive template achieves slightly better performance in the metric of landmark error. The technique detail of *hierarchical parts-based* CPD-LB is included in Appendix A. We introduce the symmetry-aware CPD in Chapter 4, which shows that symmetry-aware CPD achieves better performance in craniofacial symmetry preservation. These comparison results imply that the method described in this chapter is the best correspondence establishment framework. It may be possible to integrate the work in this chapter with our symmetry work in Chapter 4, but this was not attempted.

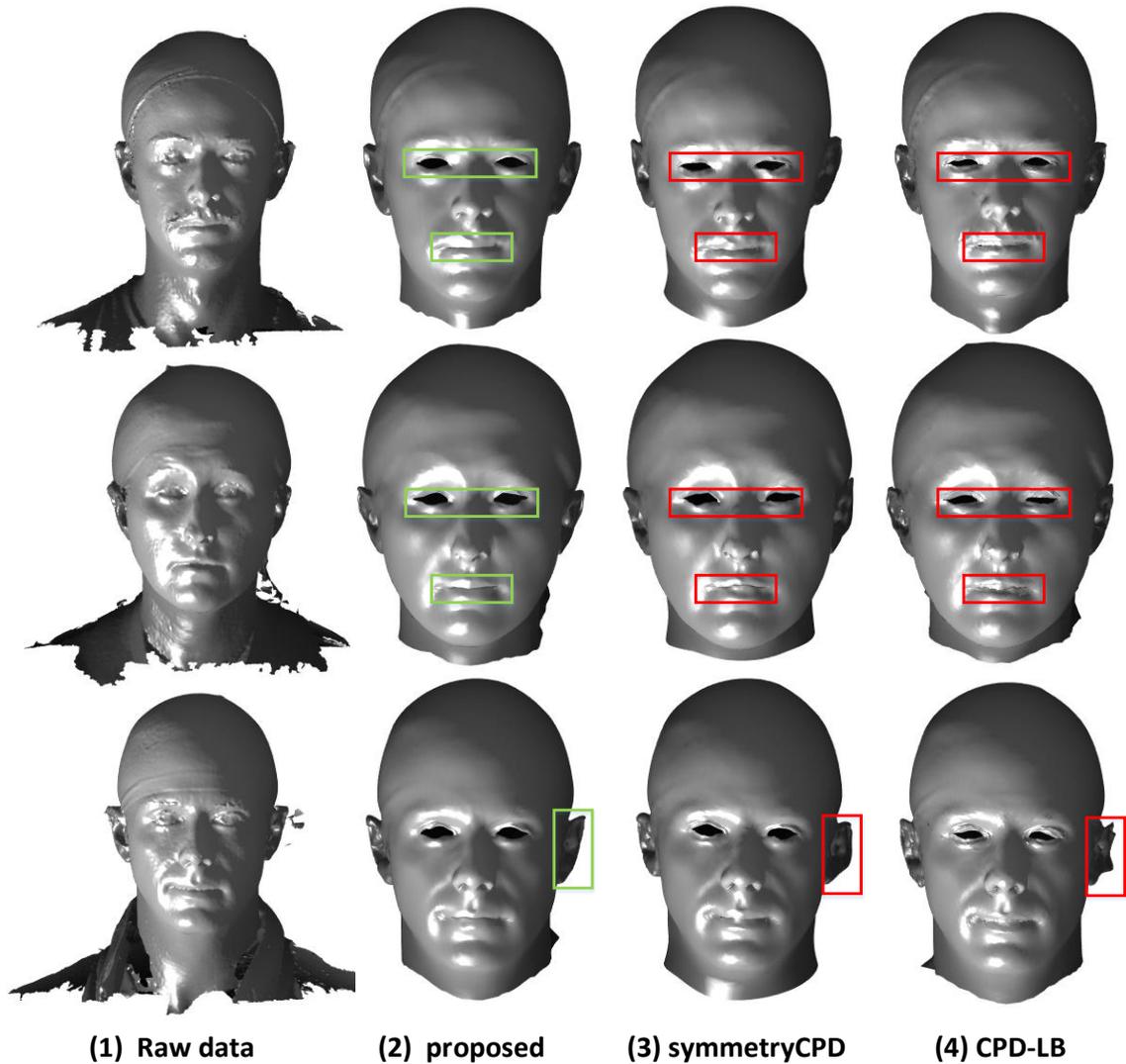


Figure 4.13: Morphing results: (1) raw mesh; (2) ICPD with adaptive template (this chapter); (3) symmetry-aware CPD (Chapter 4); (4) *hierarchical parts-based* CPD-LB (Appendix B).

### 4.6.3 Comparison with Existing Literature

In this section, we compare the proposed method with two publicly-available registration frameworks, namely the Large-Scale Face Model (LSFM) pipeline [27] and Open Framework (OF) [67], using the Headspace dataset.

**Qualitative Evaluation:** Figure 4.15 shows a typical example where the proposed method is qualitatively superior to other methods with respect to shape detail capture, and the accuracy of the mouth region. The LSFM pipeline captures shape detail but the

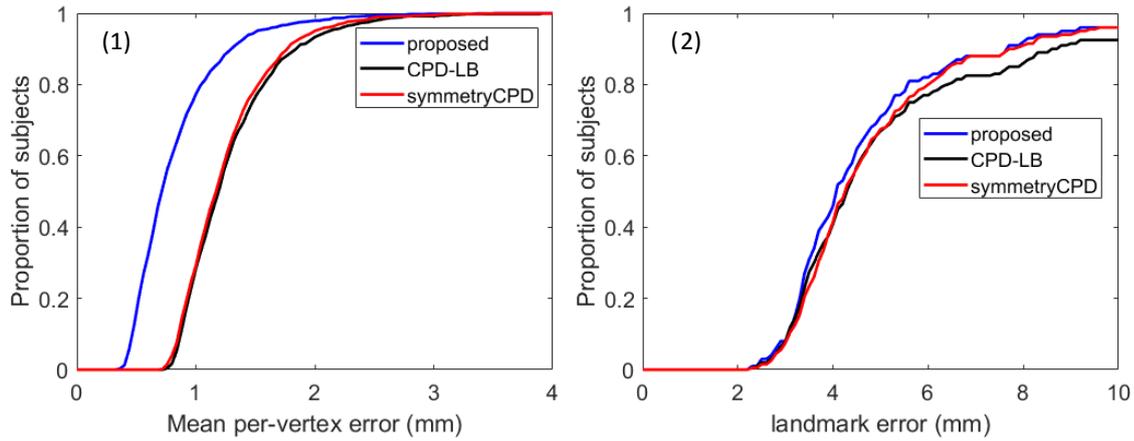


Figure 4.14: (1) Mean per-vertex nearest point error; (2) Landmark error.

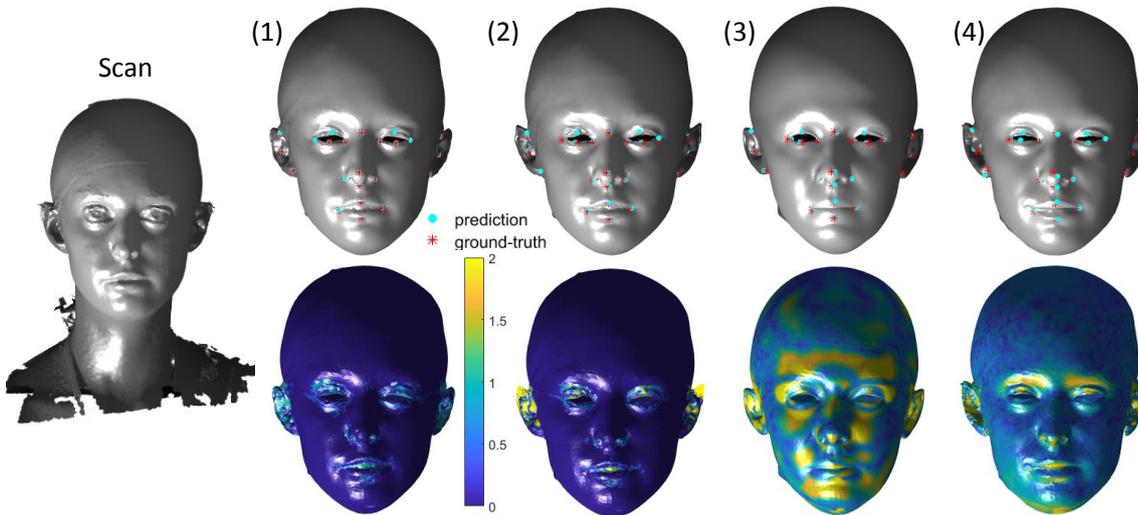


Figure 4.15: First row - correspondence results and their landmarks compared with ground-truth on raw scan; Second row - the color map of per-vertex nearest point error. (1) proposed method with LB template adaptation; (2) proposed method without adaptive template; (3) Open Framework morphing [94]; (4) LSFM morphing [26].

mouth region is not close to the raw scan. The OF pipeline has a smooth deformation field, thereby failing to capture shape detail. The OF approach requires a point projection stage after the shape registration to eliminate the shape normal error.

**Quantitative Evaluation:** Figure 4.16-left shows cumulative histograms for shape error, where shape error is computed by measuring the nearest point distance from the de-

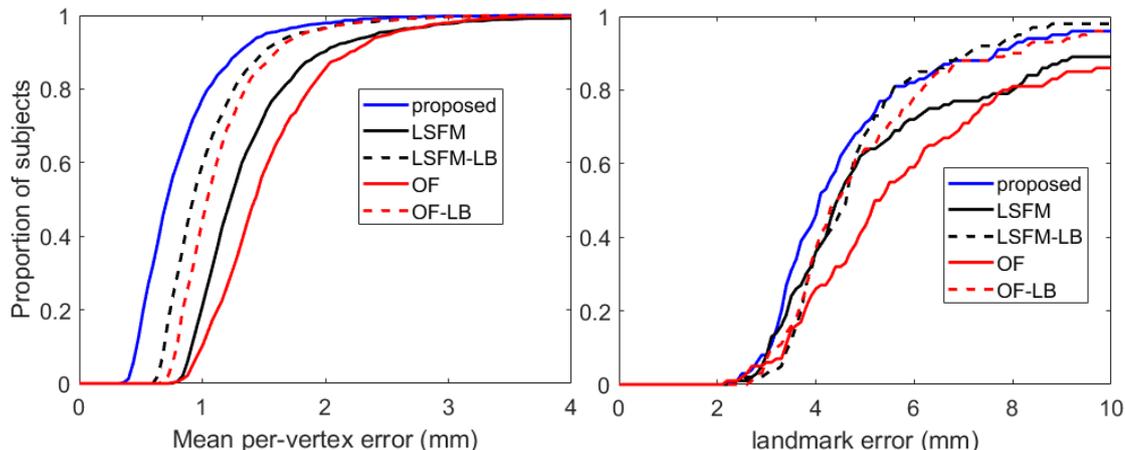


Figure 4.16: Proportion of subjects with  $<$  Euclidean distance error: left-landmark error, right-per-vertex nearest point error. The higher, the better.

formed template to the raw scan. Note that 99.12% of shape errors from our method are under 2.00mm, which is better than the other two pipelines and their extensions, which appends LB-regularised point projection to their pipelines. We use 14 manual facial landmarks over 100 subjects to measure landmark error of each template deformation method. As shown in Figure 4.16-right, 80% of landmark errors are less than 6mm for our method. Overall, the proposed method outperforms the LSFM and OF pipelines. LB-regularised point projection improves LSFM and OF in both shape error and landmark error evaluation.

## 4.7 Evaluation of 3DMMs

We select 1212 individuals (606 males and 606 females) from the Headspace dataset [1] to derive our global 3DMM by applying our fully-automatic 3DMM construction pipeline. Note that the full dataset contains 1519 subjects, but we exclude 307 subjects on the following grounds:

1. Poor fitting of the latex cap (not pulled on tightly enough).
2. Excessive hair bulge under the latex cap.
3. Excessive noise or missing parts in the 3D image.

Table 4.1: Comparison of 3DMM construction pipelines

	Initialisation	Dense correspondence	Alignment	Modelling
LSFM	Automatic facial landmarks	NICP with error pruning	GPA	PCA
OF	Manual landmarks needed	GPMM registration	GPA	GP
proposed	Automatic pose normalisation	ICPD	GPA	WPCA

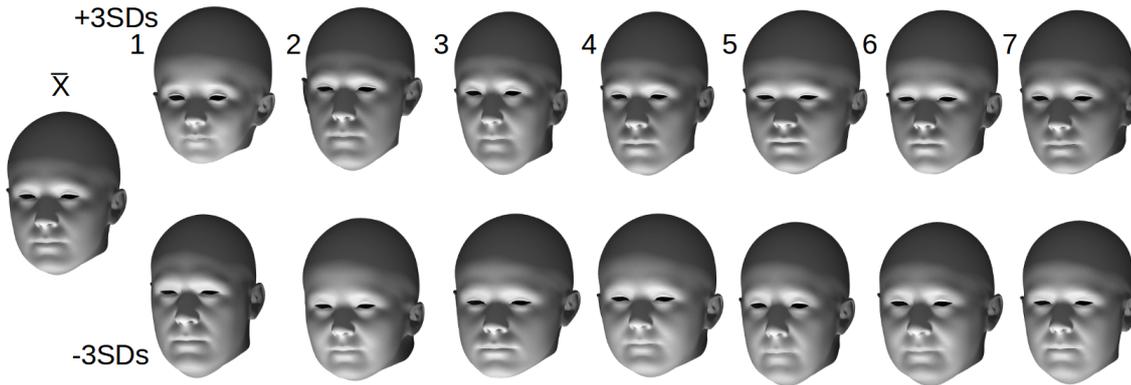


Figure 4.17: The global model of craniofacial shape: the mean and most significant 7 shape components are shown at +3SDs (top row) and -3SDs (bottom row).

4. Declared craniofacial conditions or historical head trauma.

5. A desire to balance the number of males and females in the global model build.

Subpopulations are employed to build gender-specific models, LYHM-male, LYHM-female, and four age-specific models (LYHM-age-X), focusing on demographic-specific shape variation. Section 4.7.1 is used for the visualisation of the proposed global models and sub-models on gender and age. Quantitative evaluation is presented in section 4.7.2. The performance of our 3DMMs is compared with both the LSFM pipeline [27] and Basel Open Framework (OF) [67] in section 4.7.3. Table. 4.1 shows the difference in techniques employed in pipelines. Section 4.7.4 is used for the description of shape reconstruction from an incomplete mesh. Section 4.7.5 is used for texture map image quality assessment using standard criteria.

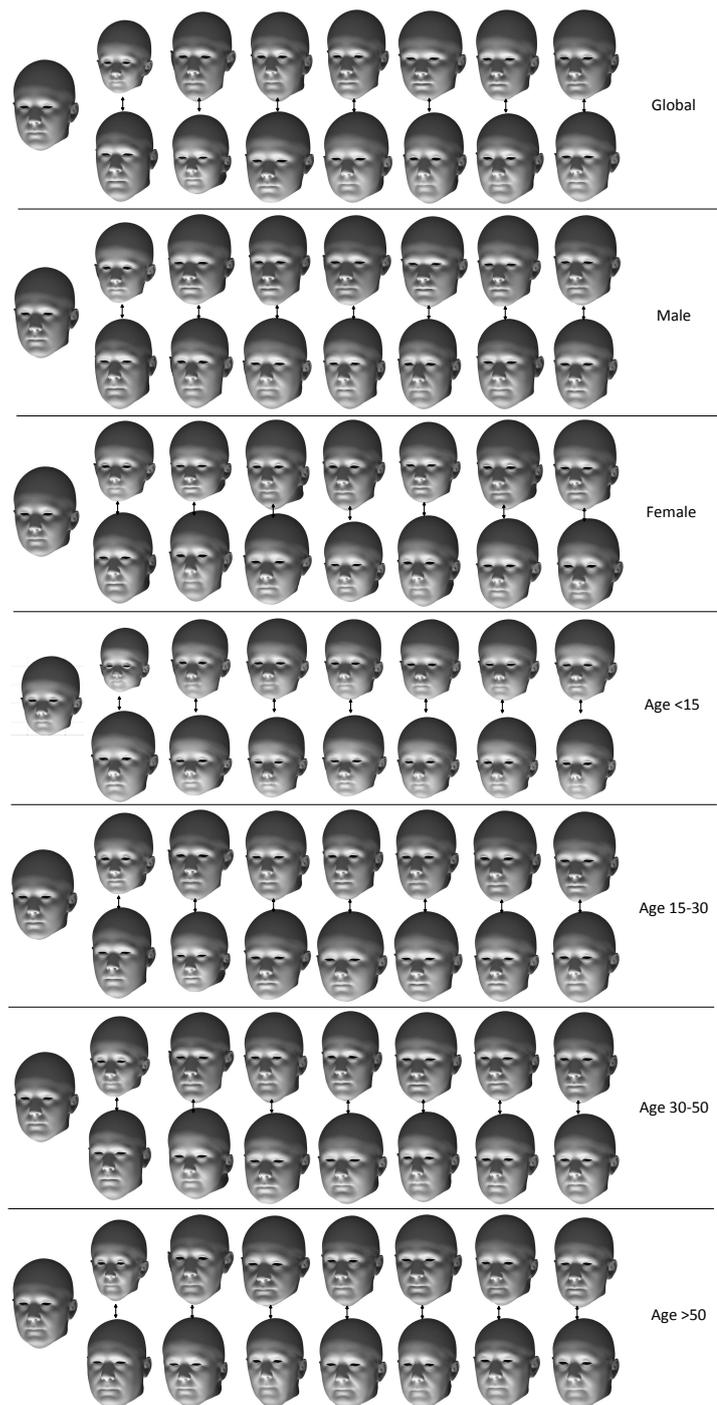


Figure 4.18: Shape models: 1st block - global model; 2nd-3rd block - gender-specific models (male and female); 4th-7th block - four age-specific models (< 15, 15-30, 30-50, > 50). In each block, the mean and most significant 7 shape components are shown at +3SDs (top row) and -3SDs (bottom row).

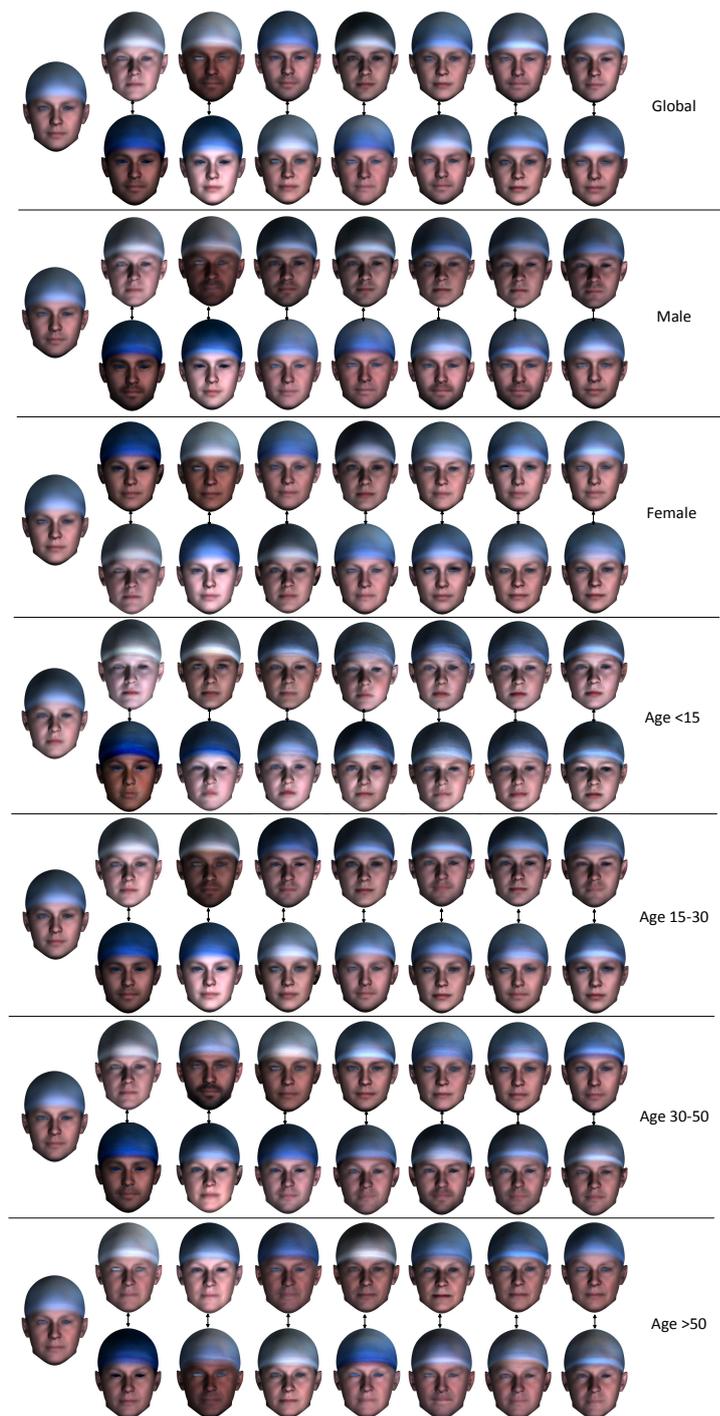


Figure 4.19: Texture models: 1st block - global model; 2nd-3rd block - gender-specific models (male and female); 4th-7th block - four age-specific models (< 15, 15-30, 30-50, > 50). In each block, the mean and most significant 7 texture components are shown at +3SDs (top row) and -3SDs (bottom row).

### 4.7.1 3DMM Visualisation

We present visualisations that provide insight into how different regions of the high-dimensional space of human face/head shape and texture are naturally related to different demographic characteristics. Figure 4.17 demonstrates the most significant modes of the global model of craniofacial shape. From  $+3SDs$  to  $-3SDs$ , the first mode shows a child head turns into an adult head, and the second mode shows a male head turns into a female head. The third mode demonstrates an obvious shape variation in jaw region. Since the craniofacial asymmetric variation is so small, it is not captured in the main modes of shape variation extracted by PCA. So, in the most important modes of the global PCA model, we cannot observe asymmetric shape variation. Taking also into account the demographics of the training data available, we define the following groups: male (all ages) and female (all ages). The dataset is further clustered into four age groups: under 15 years old, 15 – 30 years old, 30 – 50 years old and over 50 years old. The mean and most significant 7 shape components of the 6 demographic-specific models are given in Figure 4.18. Likewise, Figure 4.19 shows the mean and most significant 7 texture components of the six demographic-specific models visualized on the mean shape. The shape and texture is varied from  $+3SDs$  to  $-3SDs$ .

We apply t-Distributed Stochastic Neighbor Embedding (t-SNE) [134] to the shape components from all training samples to visualise the manifold of high-dimensional training shapes. By applying t-SNE between the shape parameters and its corresponding demographic data in our dataset, we are able to label samples in this space by their gender, see Figure 4.20 (left). We observe that t-SNE has produced a nonlinear 2D embedding that emphasises gender difference. Figure 4.20 (right) shows the visualisation of the space by age. Interestingly, a clear trend of increasing age can be seen across the space, suggesting that the facial shape space has age-related structure.

### 4.7.2 Evaluation of Sub-models

According to the demographic data in our dataset, we build sub-models based on gender and age. For quantitative model evaluation, Styner et al [123] give detailed descriptions of three metrics: compactness, generalisation and specificity. The compactness of the model

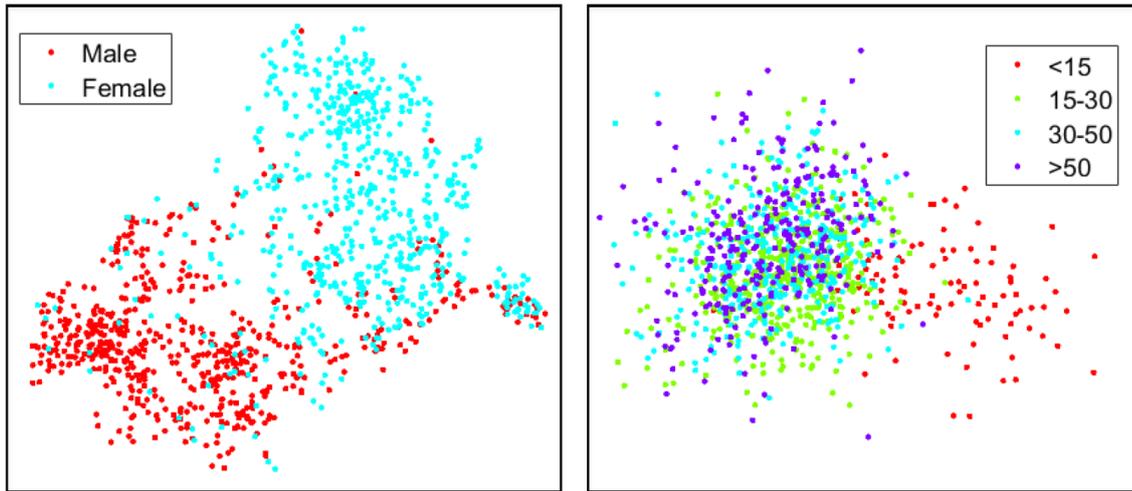


Figure 4.20: t-Distributed Stochastic Neighbor Embedding in terms of gender (left) and age (right) groups.

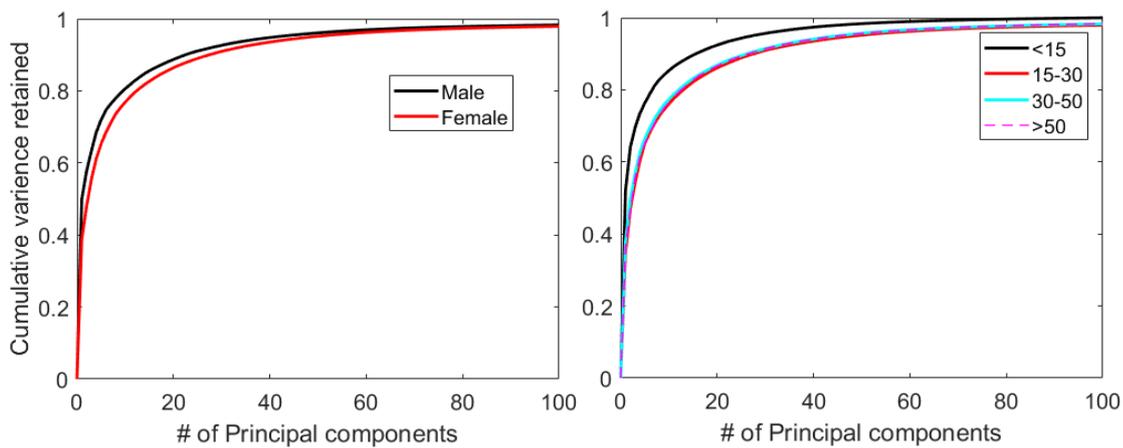


Figure 4.21: Compactness of sub-models.

describes the number of parameters required to express some fraction of the variance in the training set, fewer is better. As can be seen from Figure 4.21, LYHM-male and LYHM (<15) have better compactness than other demographic specific models. In Figure 4.22, LYHM-female and LYHM (15-30) have the lower generalisation error, which implies that LYHM-female and LYHM (15-30) have better performance in describing unseen examples. Specificity measures how well a model is able to generate instances that are similar to real data. Figure 4.23 shows LYHM-male and LYHM (15-30) have lower distance error, which implies that LYHM-male and LYHM (15-30) is better at generating instances

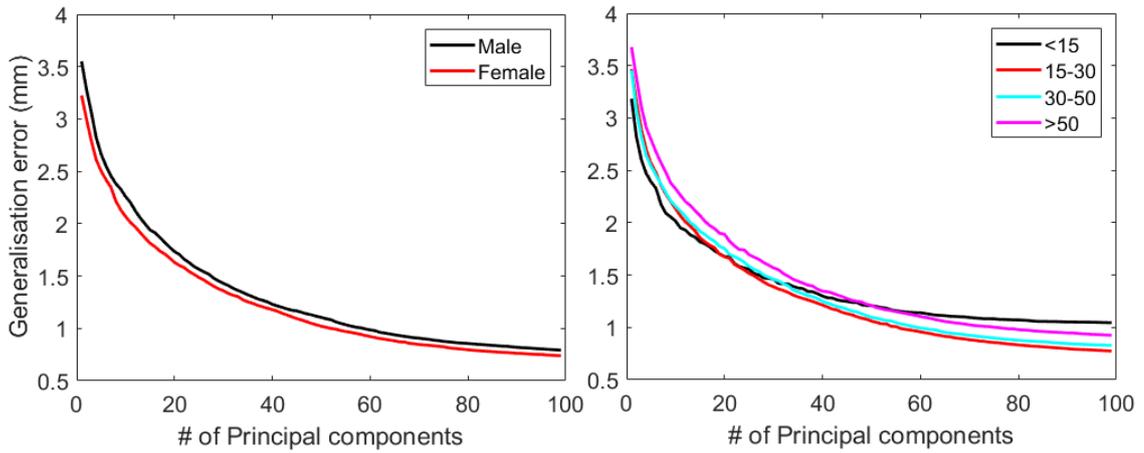


Figure 4.22: Generalisation of sub-models.

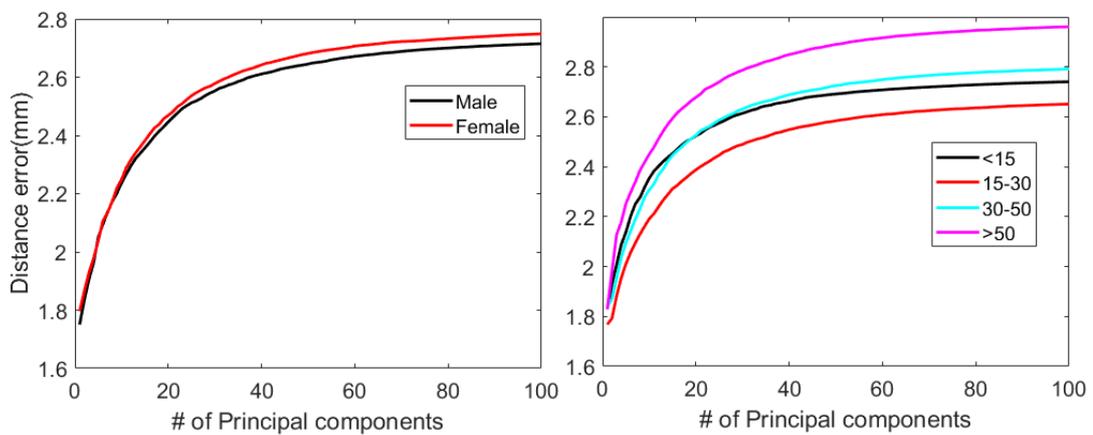


Figure 4.23: Specificity of sub-models.

close to real data.

### 4.7.3 Evaluation of Full Head 3DMMs using 3DMM Pipelines in the Literature

We build full head 3DMMs using the proposed method, the LSFM pipeline [27], and the OF pipeline [67], again with 1212 subjects in Headspace dataset. The compactness of the model describes the number of parameters required to express some fraction of the variance in the training set, fewer is better. We use the same subjects as LYHM-global to build 3DMMs. As can be from Figure 4.24 (a), when less than 33 components

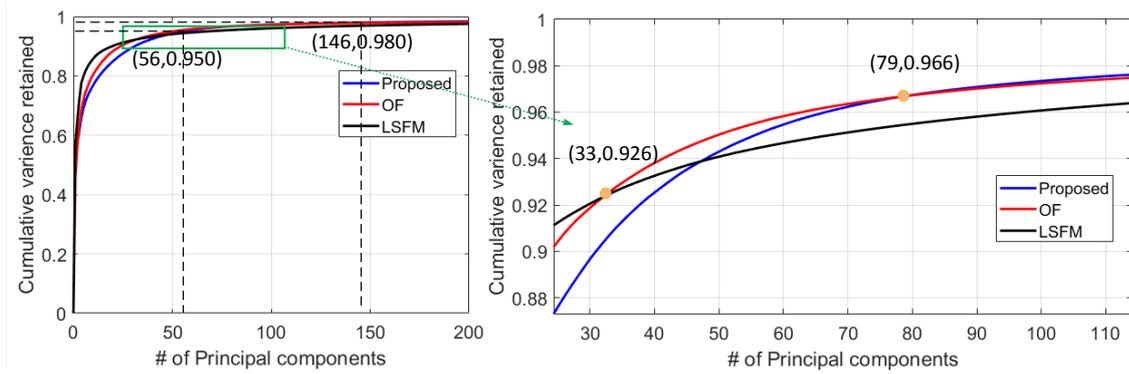


Figure 4.24: Compactness, higher is better.

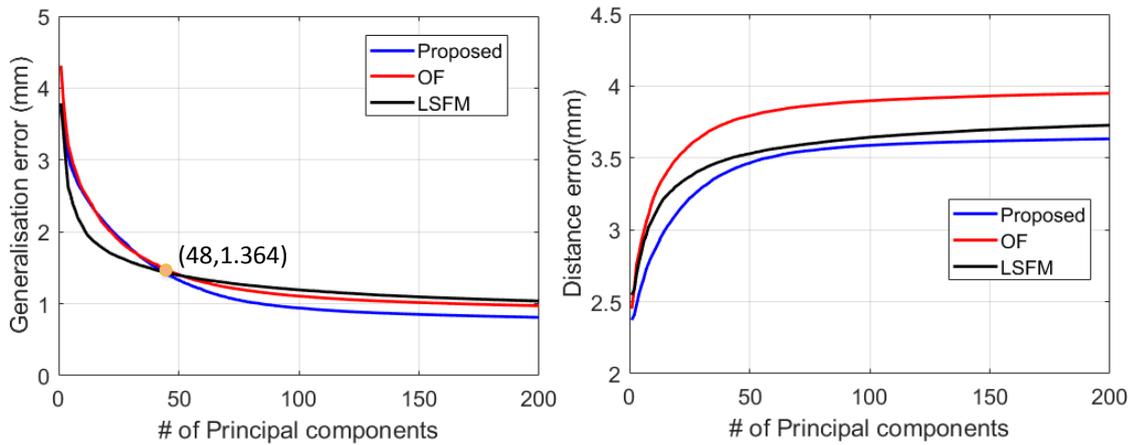


Figure 4.25: Generalisation (left), lower is better; Specificity (right), lower is better.

used, LSFM is more compact than the proposed method and OF. Between the first 33 and 79 components, the model constructed by OF is more compact than the other two. When more than 79 components used, the proposed method has better compactness than LSFM [27] and OF [67]. With the first 56 and the first 146 components used, the 3DMM constructed by the proposed method retains 95% and 98% of the shape variation in the training set.

The generalisation of the model demonstrates the ability in generating unseen examples. With fewer than 48 model parameters, LSFM is better than the proposed method in terms of generalisation error. With more than 48 components, the proposed method has the lowest generalisation error, which in this case implies that proposed method has the best performance in describing unseen examples.

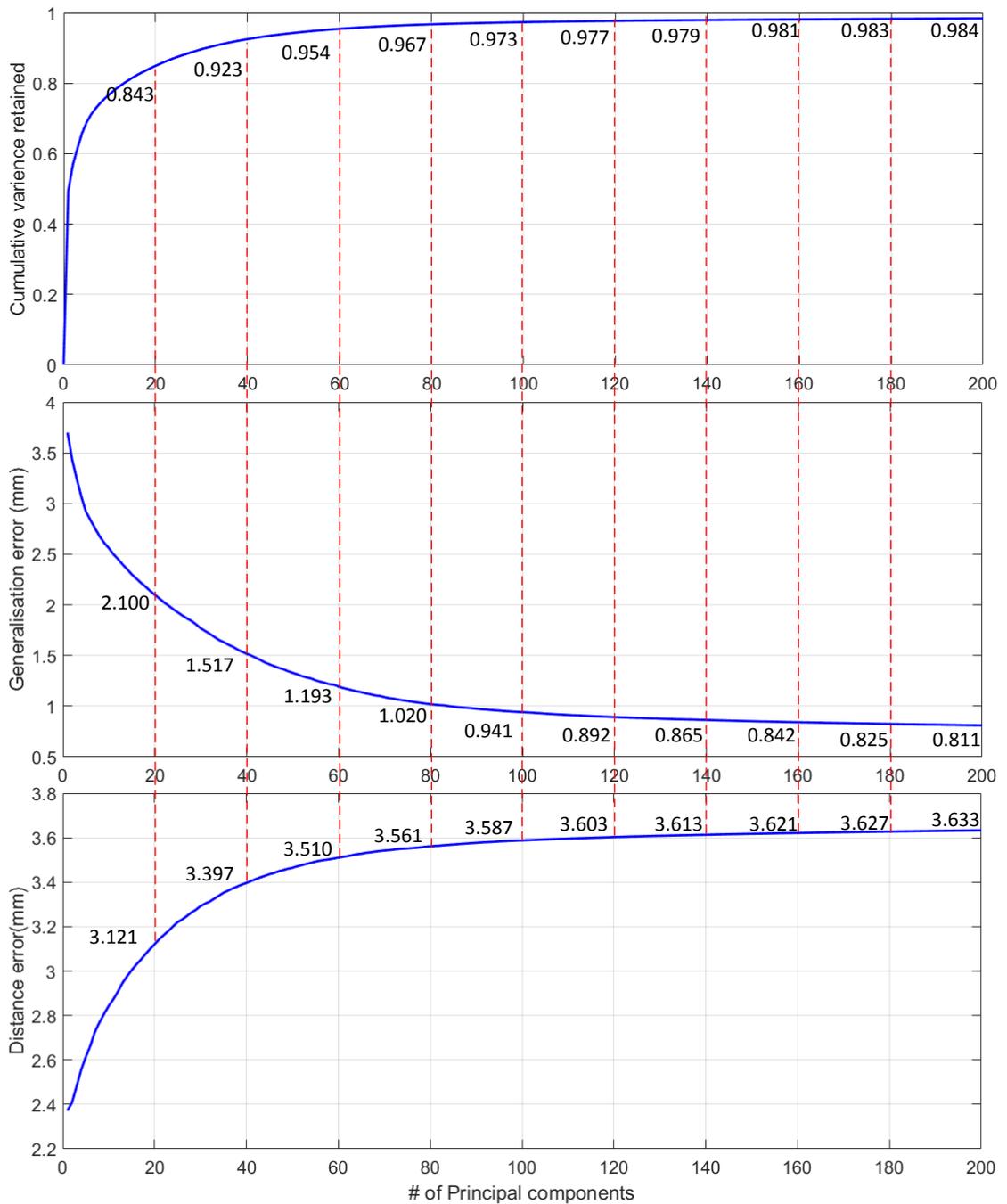


Figure 4.26: Critical analysis of the proposed model in terms of the number of model components.

Specificity measures how well a model is able to generate instances that are similar to real data. The proposed method has the lowest distance error, which implies that the proposed method is best at generating instances close to real data.

Overall, with more 79 components used, the proposed pipeline is better than LSFM [27] and OF [67] in terms of the compactness of model. The generalisation error of LSFM decreases faster than the proposed method. But with more components used, the proposed method has the lowest generalisation error when compared with the other two pipelines. The proposed method outperforms LSFM and OF in specificity.

**Critical Analysis of the Number of 3DMM Components:** It is desirable to have a lower generalisation and specificity error for a 3DMM, but there is a trade-off between generalisation and specificity. The compactness is also important to ensure that most of shape variation is captured by the model using a relatively small number of model parameters. So we now perform a critical analysis of the number of components that should be used for 3DMM. The actual measurements are provided in Figure 4.26. When the application requires that the error of describing unseen examples should be less than 1.00 mm, the distance error from random sample to real data should be less than 3.60 mm, and the model should capture more than 97% of the shape variation, the best choice is the first 100 components to be used in the 3DMM. In general, when more than 80 components are used in the 3DMM, the specificity error and compactness curves flatten (i.e. they increase slower and slower), also the generalisation curve flattens (i.e. the errors decrease slower and slower). With the first 80 components used in the 3DMM, 96.7% of the shape variation is captured, the error in describing unseen examples is 1.02 mm and the distance error from random sample to real data is 3.56 mm. As discussed in the evaluation results, with more than 79 components used, the model constructed by the proposed method has the best performance when compared with LSFM and OF across all three metrics: compactness, generalisation and specificity. With more components used, more noisy shape variation is introduced and more computation is required. Taking all these into consideration, in this chapter we choose to employ the first 100 model components for our applications.

#### 4.7.4 Reconstruction

The task of reconstruction of an unseen example with missing parts can validate the correspondence quality, model fitting and generalisation ability [3]. In the first row of Figure 4.27, we fix the shape of the cranium and reconstruct the full head from that shape, while

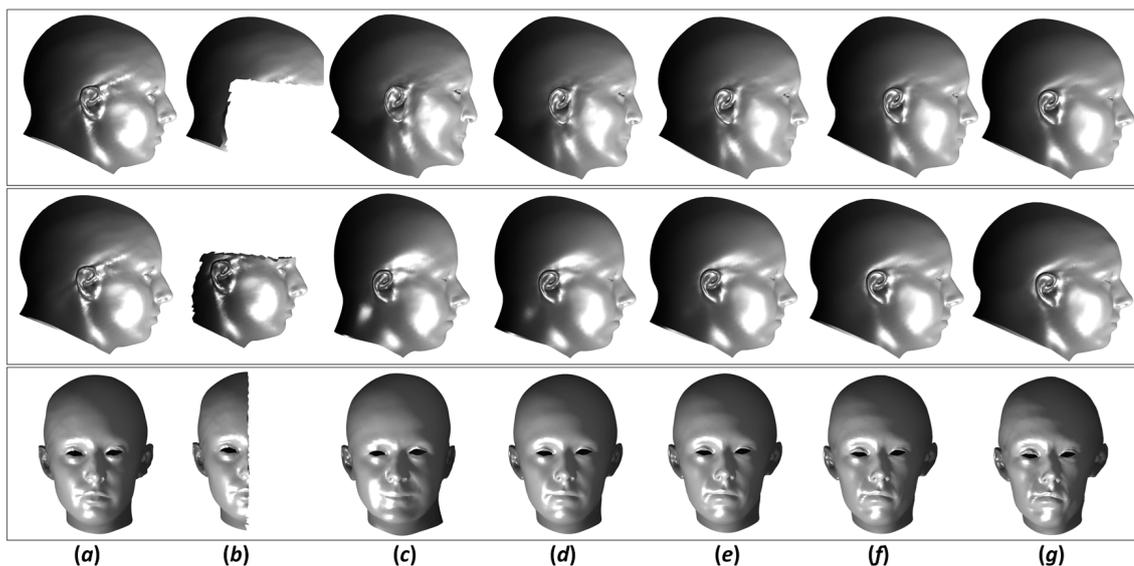


Figure 4.27: The flexibility of reconstruction from incomplete data: (a) example; (b) crop (incomplete data); (c)-(g) are the flexibility modes, and (e) is the mean.

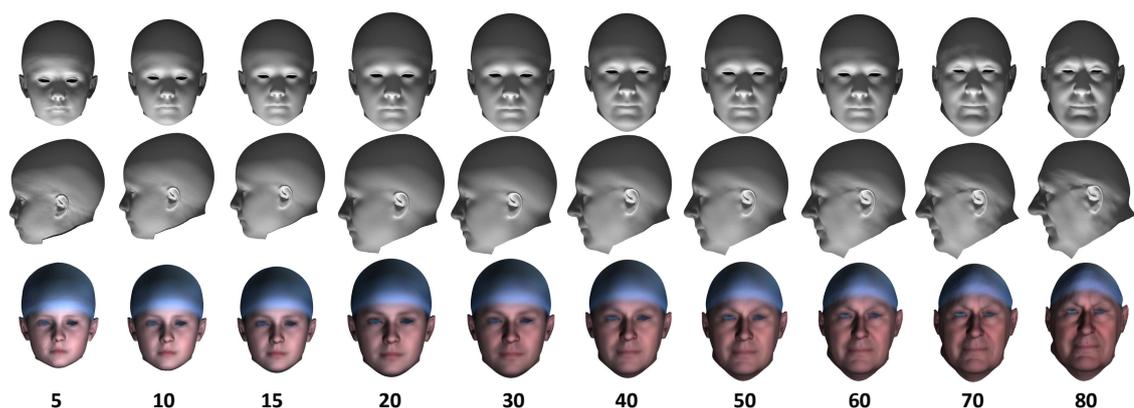


Figure 4.28: Age regression from 5 years to 80 years.

permitting the shape of the face to vary. Here we found that most variation occurs over the chin region. The second row of Figure 4.27 is the reconstruction from the face only, and we note that the principal variation in reconstructed cranium is from high/short to low/long. This offers a way to augment existing face models with a cranium. Reconstruction from one side of the sagittal symmetry plane demonstrates asymmetrical variation of the head. This application can aid shape-from-shading using 2D profile images to predict the 3D shape of the self-occluded half of the face.

Table 4.2: Texture map image quality assessment

	SSIM	MS-SSIM	IW-SSIM
Per-vertex color	0.8790	0.8618	0.6238
Texture mapping	<b>0.8926</b>	<b>0.8712</b>	<b>0.6505</b>

### 4.7.5 Texture Map Image Quality Assessment

As shown in Figure 4.10, the proposed texture map technique outperforms per-vertex texture image qualitatively. We use several metrics, namely: SSIM [138], MS-SSIM [140], and IW-SSIM [139] to measure the texture map quantitatively. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, Structural SIMilarity (SSIM) is based on the degradation of structural information (higher is better). We save the rendering of the raw image, the proposed texture map and the per-vertex texture image into the same image size. When using the full-reference image quality assessment indices, we treat the rendering of the raw image as the reference image for image quality assessment. The rendering of the proposed texture map and per-vertex texture image are compared with this reference image, respectively. As can be seen in Table. 4.2, the proposed texture mapping technique improves the texture image quality for texture modelling, when compared with that of per-vertex color texture image.

## 4.8 Applications

We now demonstrate three applications of 3DMMs, including: age regression in section 4.8.1, clinical intervention outcome evaluation in section 4.8.2 and craniofacial anthropometric measurements in section 4.8.3.

### 4.8.1 Age Regression

Model regression can validate the model's generalisation ability on some property (e.g. age, gender). We use a multiple linear regression function *regress* in Matlab between the shape-and-texture parameters against age. This function returns a vector of coefficient estimates for a multiple linear regression of the responses in shape-and-texture parameters

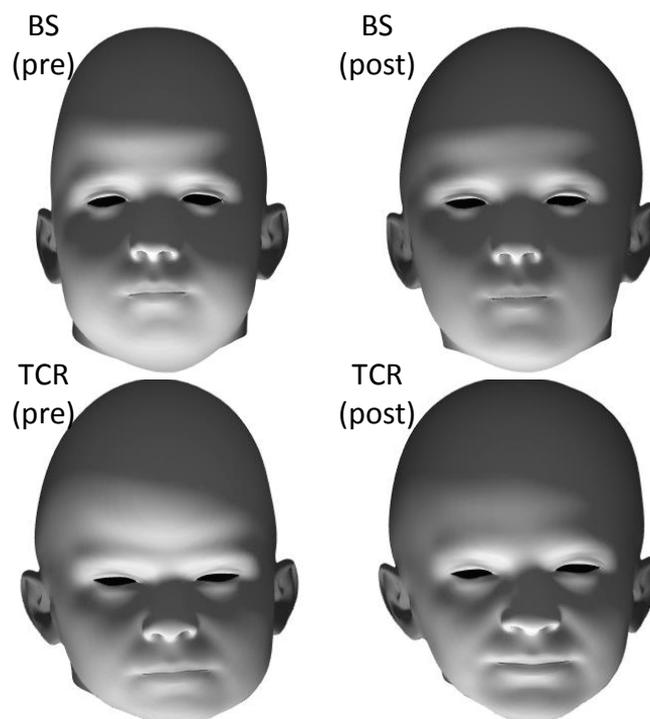


Figure 4.29: 3D meshes of patients (pre-operation and post-operation) after registration.

on the predictors in the matrix of age. After this learning process, shape parameters can be revisited when given the age value. Then the shape for the specific age can be reconstructed by adding the shape variation to the mean shape. Figure 4.28 demonstrates a sequence of 3D images generated from age regression of the proposed model, ranging from 5 to 80 years. Note that the cranium develops before the face in children, and the shape of the cranium is not changing in adults.

### 4.8.2 Clinical Intervention Outcome Evaluation

In this section, we use one of our demographic models to describe post surgical change in a sample of 17 craniosynostosis patients (children), 10 of which have undergone one type of cranial corrective procedure *Barrel Staving* (BS) and the other 7, another cranial corrective procedure *Total Calvarial Remodelling* (TCR).

Note that both facial structure and overall scale are now irrelevant and that major cranial shape changes are not thought to occur after 2 years old. The LYHM (<15, cranium) model is applied to intervention outcome evaluation and its mean is used as the template.

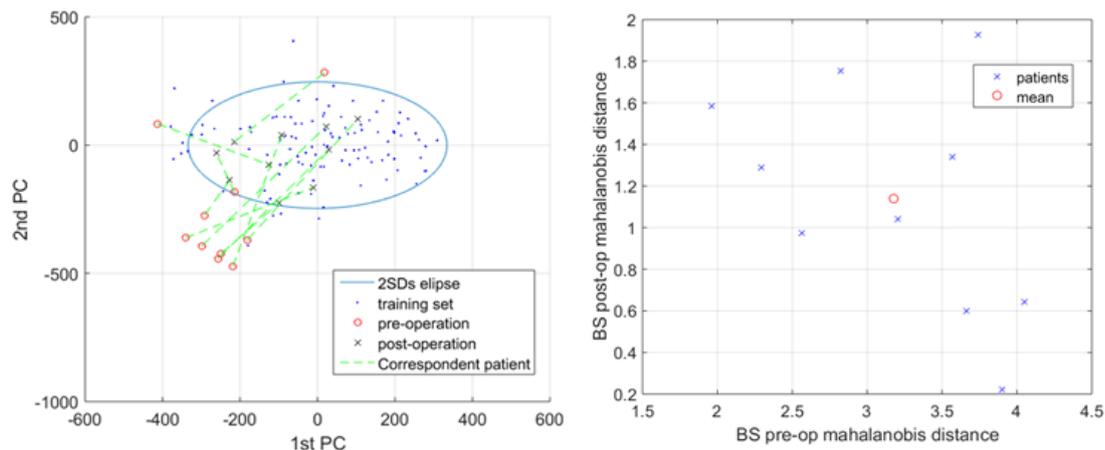


Figure 4.30: BS intervention outcome: left - first two principal components of patients compared to general population; right - pre-operation against post-operation in terms of Mahalanobis distance of first two principal components of patients.

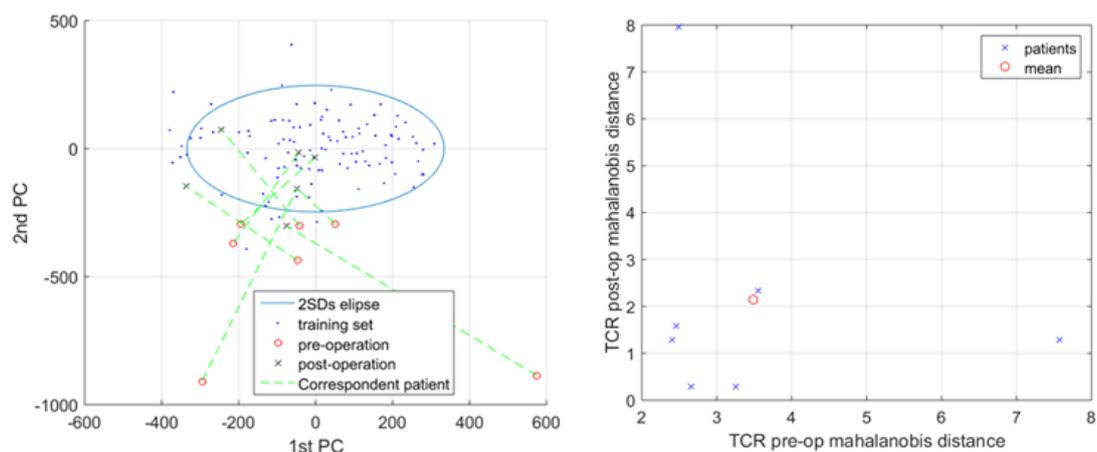


Figure 4.31: TCR intervention outcome: left - first two principal components of patients compared to general population; right - pre-operation against post-operation in terms of Mahalanobis distance of first two principal components of patients.

We treat the patients as the unseen example for LYHM (<15, cranium). We can then plot their pre-operative and post-operative parametrisations and compare them with the parameterisations of LYHM (<15, cranium). The expected result is that the parameterisations should show the head shapes moving nearer to the mean of the training examples. Figure 4.29 demonstrates the 3D meshes of patients (pre-operation and post-operation) after registration. The results are shown in Figure 4.30 and Figure 4.31. The parameter-

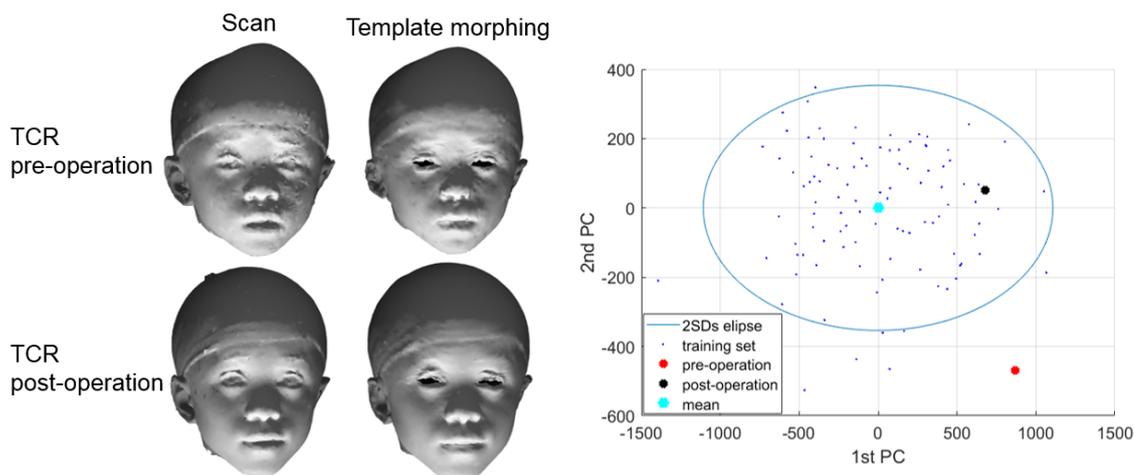


Figure 4.32: A case study on pre-operation and post-operation of a specific patient in terms of full head shape.

isations are validated to move nearer to the mean, which is at the origin of the plots. To our knowledge, this is the first use of full head 3DMMs in a craniofacial clinical study.

Figure 4.32 demonstrates a case study on pre-operation and post-operation of a specific patient. The surgeons are also interested in the influence of operation on facial shape. Here, we use the shape of full head, both face and cranium, for analysis. We can clearly observe the improvement after operation when viewing the 3D shape. This is validated by the shape analysis. The shape parameters of pre-operation is outside 2 Standard Deviations (SDs) ellipse of the training set, while the shape parameters of post-operation is within 2 Standard Deviations (SDs) ellipse.

To evaluate it quantitatively, we calculate the Mahalanobis distance of each patient. As can be seen in Figure 4.30-right and Figure 4.31-right, the mean of Mahalanobis distance for all patients decreases from 3.21 to 1.18 standard deviations in terms of the BS operation. For the TCR operation, the mean of Mahalanobis distance for all patients decreases from 3.52 to 2.23 standard deviations. The improvement with BS is 63.24 % and the improvement with TCR is 36.65 %. In our case study, BS has better outcomes of clinical intervention than TCR, but the population sample is too small to be conclusive.

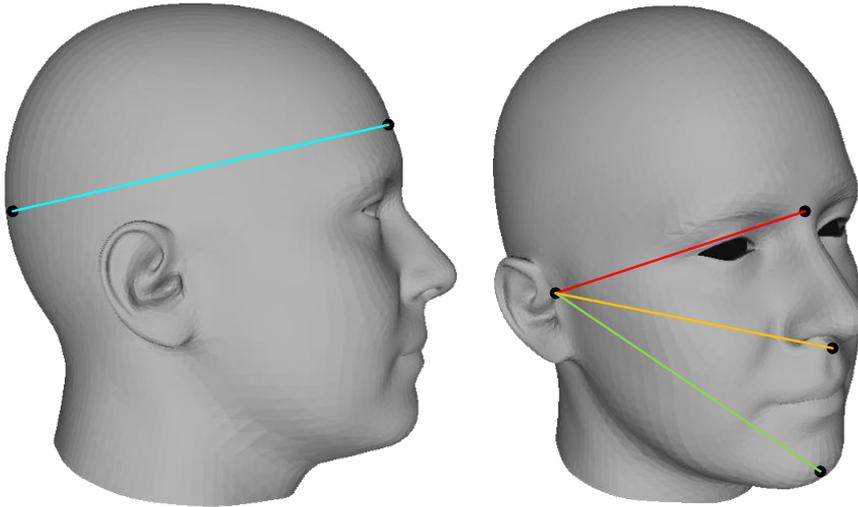


Figure 4.33: Craniofacial measurements: left - Maximum Cranial Length; right - Upper Facial Depth (red line), Middle Facial Depth (yellow line) and Lower Facial Depth (green line).

### 4.8.3 Craniofacial Anthropometric Measurements

Craniofacial anthropometric measurements are very useful for plastic surgery. The surgeons usually use many measuring instruments to do the measurements directly on a patient's face. As shown in Figure 4.33, four types of length are provided for the measurements of maximum cranial length, upper facial depth, middle facial depth and lower facial depth. Once these six landmarks are clicked on the template mesh, the indices of these landmarks are known for calculation of the four types of length.

With the proposed method, we provide a fully automatic way to do facial anthropometric measurements with 3D scanners. After capturing the 3D shape of one patient, we use the proposed template morphing framework to build dense correspondences. Since the morphed template has the same indices of these landmarks as the template mesh. Then we use these indices to search the 3D axis coordinates of these landmarks on the morphed template.

We can expect a significant growth of the face and cranium during youth and a smaller growth in old age. This is validated in Figures 4.34 and 4.35. The noisiness of the measurements in old age is due to fewer samples. We compare the measurements in Headspace for 1212 subjects with those in FaceBase [117]. In both datasets, we can

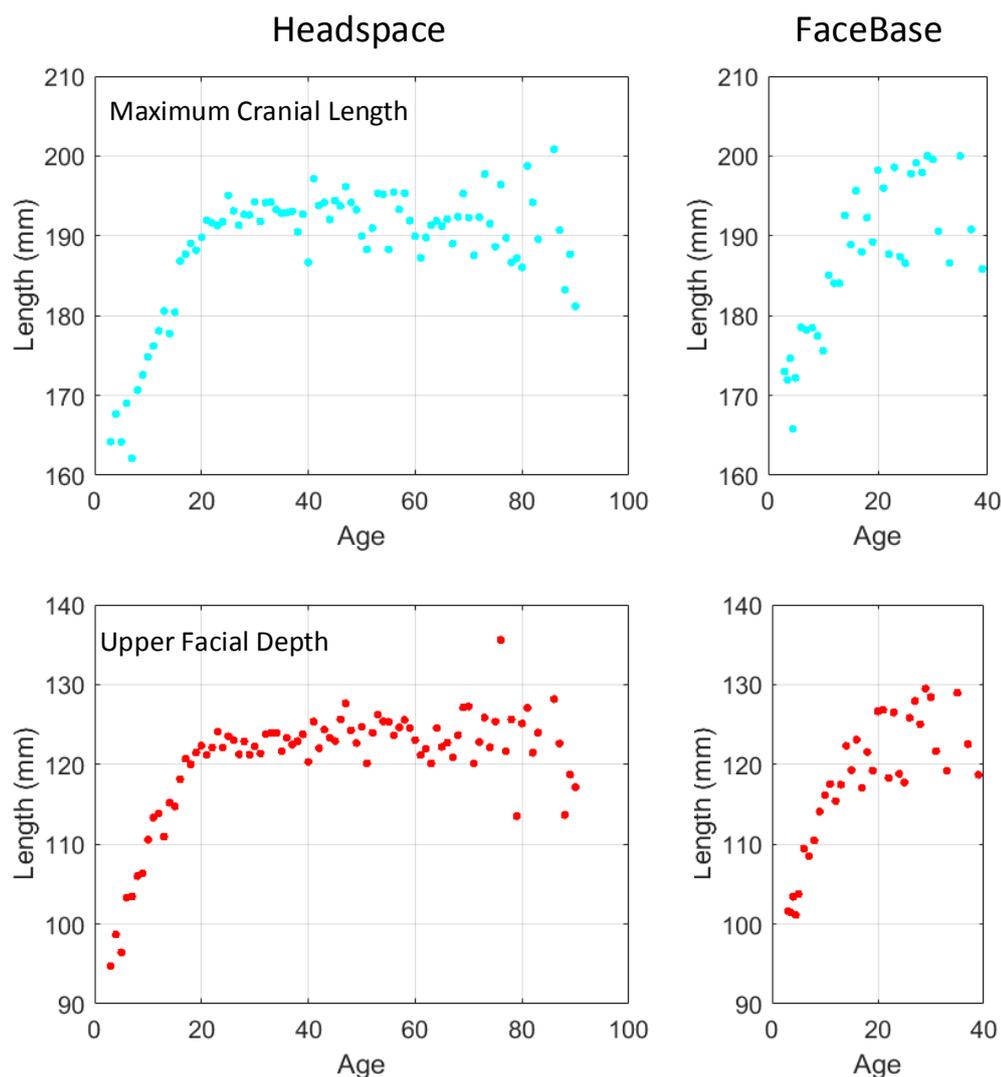


Figure 4.34: Craniofacial Anthropometric Measurements over Headspace (right) and FaceBase (left) dataset: 1st row - Maximum Cranial Length and 2nd row - Upper Facial Depth.

observe that the face and cranium grows significantly before 20 years old. After that, the growth is very small and stable. In Headspace, we can see the motion of face and cranium growth after 40 years old. A small growth can be observed after 40 years old. We can chose a starting point  $M_s$  and an ending point  $M_e$ . We can calculate the growth rate of face and cranium from 0 to 20 years old:

$$\text{Gr} = \frac{(M_e - M_s)}{M_s} \quad (4.14)$$

where  $\text{Gr}$  indicates the growth-rate. The growth rate for maximum cranial length, up-

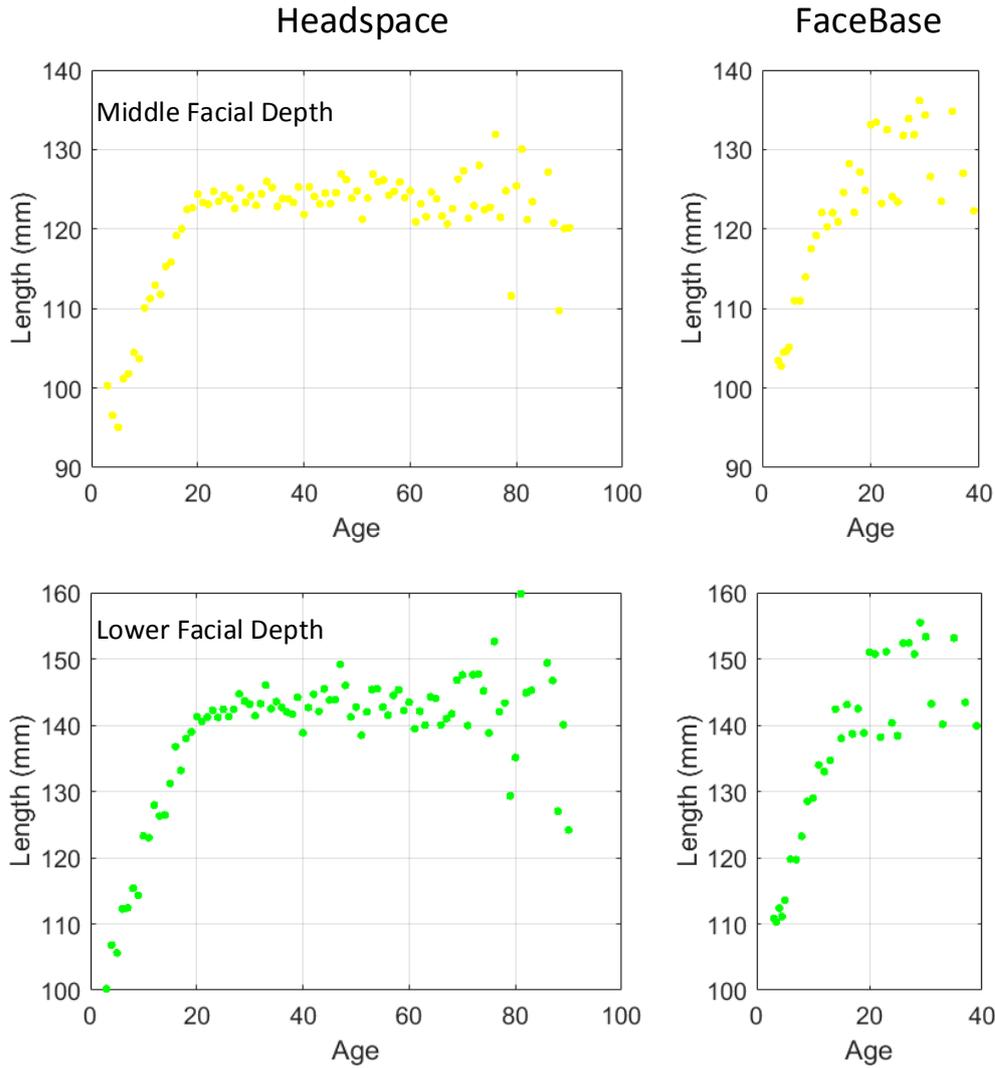


Figure 4.35: Craniofacial Anthropometric Measurements over Headspace (right) and FaceBase (left) dataset: 1st row - Middle Facial Depth and 2nd row - Lower Facial Depth.

per facial depth, middle facial depth and lower facial depth are about 18.75%, 33.33%, 33.33% and 40.00%. The face grows more significantly than the cranium. The lower facial depth has the most significant growth when compared with the other three measurements.

## 4.9 Summary

We released the first publicly-available full head dataset with meta-data for academic research. We proposed a fully-automatic 3DMM training pipeline and used it to build

the first shape-texture 3DMM of the full head. The correspondence framework avoids over-fitting and under-fitting in template morphing. The adaptive template improves the correspondence accuracy in local regions, while the LB regularised point projection helps in decreasing the correspondence error in the normal direction. The correspondence accuracy is state-of-the-art, in terms of publicly-available pipelines. The texture mapping technique captures high quality texture for texture modelling. The proposed 3DMMs have a powerful ability in reconstruction of incomplete data and model regression to observe the influence of age on craniofacial growth. The flexibility of reconstruction from incomplete craniofacial data helps in many computer vision applications. We present the first use of statistical 3D craniofacial shape models in a clinical study. Both BS and TCR operations show a shape improvement due to clinical intervention. We provided a fully automatic way to do facial anthropometric measurements.



## Chapter 5

# Symmetric Morphing and Symmetry-factored Statistical Modelling

Most biological objects, including human faces and heads, possess approximate symmetries. Often this is principally extrinsic, bilateral symmetry (i.e. reflective symmetry about a plane bisecting the object). Deviations from exact symmetry are an interesting and potentially important geometric property in terms of modelling and understanding 3D shape variation. The degree of asymmetry may convey information about an object. For example, there is evidence that facial asymmetry is used by humans to measure genetic health [81] and asymmetry in man-made objects may indicate imperfections in the manufacturing process. For this reason, we suggest constructing statistical shape models that separate symmetric from asymmetric shape variability, in order to reveal interesting and potentially subtle aspects of shape variation over population datasets.

The asymmetric variation in human head shape is much smaller than symmetric variation. Hence, classical statistical models, obtained by applying PCA directly to aligned and normalised head shape datasets, pools asymmetric variation with the much larger symmetric variation. The asymmetric variation is thus difficult to capture in a single linear model. The decomposition of morphable models into symmetric and asymmetric variation is able to express asymmetry variation in heads explicitly, without that variation being pooled with the much larger symmetrical variation.

Symmetry-factored statistical modelling of craniofacial shape opens up new opportunities not afforded by existing 3D morphable models (3DMMs). We can analyse the asymmetry of multiple craniofacial regions. We can also investigate the importance of symmetry-factored information in different applications, e.g. face recognition or ergonomics.

As shown in Figure 5.1, our symmetry modelling framework includes two main stages: 1) symmetric deformation and 2) symmetry-factored modelling. We decompose the symmetric deformation into global symmetric deformation and local symmetric deformation, which can be seen as symmetry-plane aware CPD (SA-CPD), followed by LB-regularised point projection. The global/local symmetric deformation is to symmetrise the global/local deformation field.



Figure 5.1: Symmetry-factored statistical modelling framework.

Why do we need symmetric deformation results for symmetry-factored modelling? When extracting the symmetric and asymmetric shape variation from the registered data, we have the hypothesis that the meshes have been symmetrically remeshed. This means that for any vertex  $i$ , its symmetric partner is given by  $sym(i)$ . Vertices lying on the symmetry line are self-symmetric, i.e.  $i = sym(i)$ . As can be seen from Figure 5.2, a symmetric partner exists in a template with a perfectly symmetric face. However, this is not true when the symmetry contour slides to one side, as shown in Figure 5.2 (3). In other words, vertex  $j$  is not the symmetric partner of vertex  $i$  in this deformation result. Figure 5.2 (4) shows a typical result of the proposed symmetric deformation. The symmetry contour of the deformation result (red points) is very close to the symmetry contour of the target data (blue points).

**Pipeline Design Motivation.** Note that, as well as *motion field* symmetrisation during deformation, we use point set symmetrisation later in the pipeline. Why is this necessary? Figure 5.3 (1) demonstrates the end result of symmetry-plane aware CPD, in which the deformation field is symmetrised. The symmetry contour is forced to be fixed to the

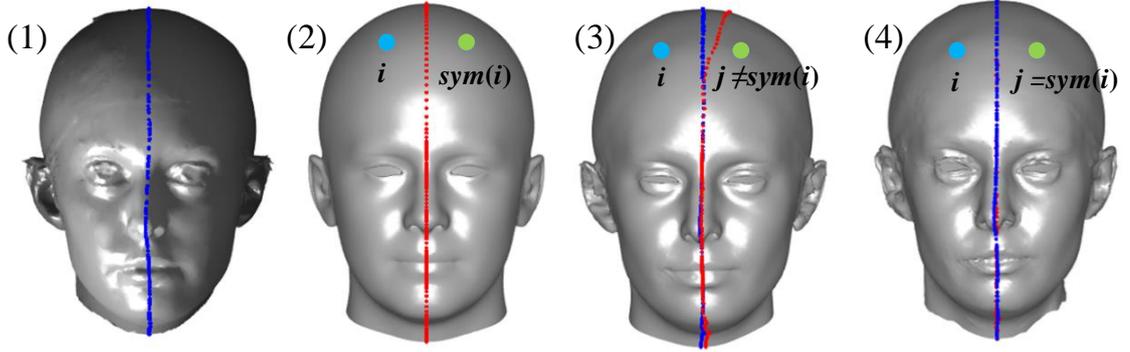


Figure 5.2: Blue points are the symmetry contour on the target data. Red points are from the symmetry contour indices of the source template: (1) scan, (2) template, (3) deformation result with no symmetry constraint, (4) proposed symmetric deformation.

$X = 0$  plane. We then use LB-regularised point projection to capture the shape detail, which breaks the morphed template symmetry. Therefore we then symmetrise the end result to generate the symmetric face for the given subject. The symmetric face for the subject can be defined as: 1) the mirror of right face is exactly the same as the left face; 2) when this symmetric face is subtracted from the end result shown in Figure 5.3 (2), only the asymmetric shape variation is left. Note that the end result of the *deformation field symmetrisation* is also symmetric, as shown in Figure 5.3 (1), but it includes other symmetric shape variation that does not belong to the specific subject. This is why point-based symmetrisation is not redundant, when deformation field (i.e motion-based) symmetrisation is performed on previous stage. Deformation field symmetrisation is required so as not to destroy the symmetry property during the deformation. Overall, the symmetric deformation aims at preserving the symmetry-plane. This ensures that vertices lying on the symmetry line are self-symmetric, which is necessary for point set symmetrisation. Point set symmetrisation aims at generating the symmetric face for the specific subject and extracts the asymmetric shape variation.

**Chapter Structure.** The remainder of this chapter is structured as follows. Section 5.1 is used for mathematical notation and preliminaries. In Section 5.2, we describe our new symmetric shape morphing process. Registered data is the output of symmetric deformation algorithms. We evaluate this in section 5.3. We use the registered data as the input of our symmetry-factored modelling. We present our symmetry-factored modelling

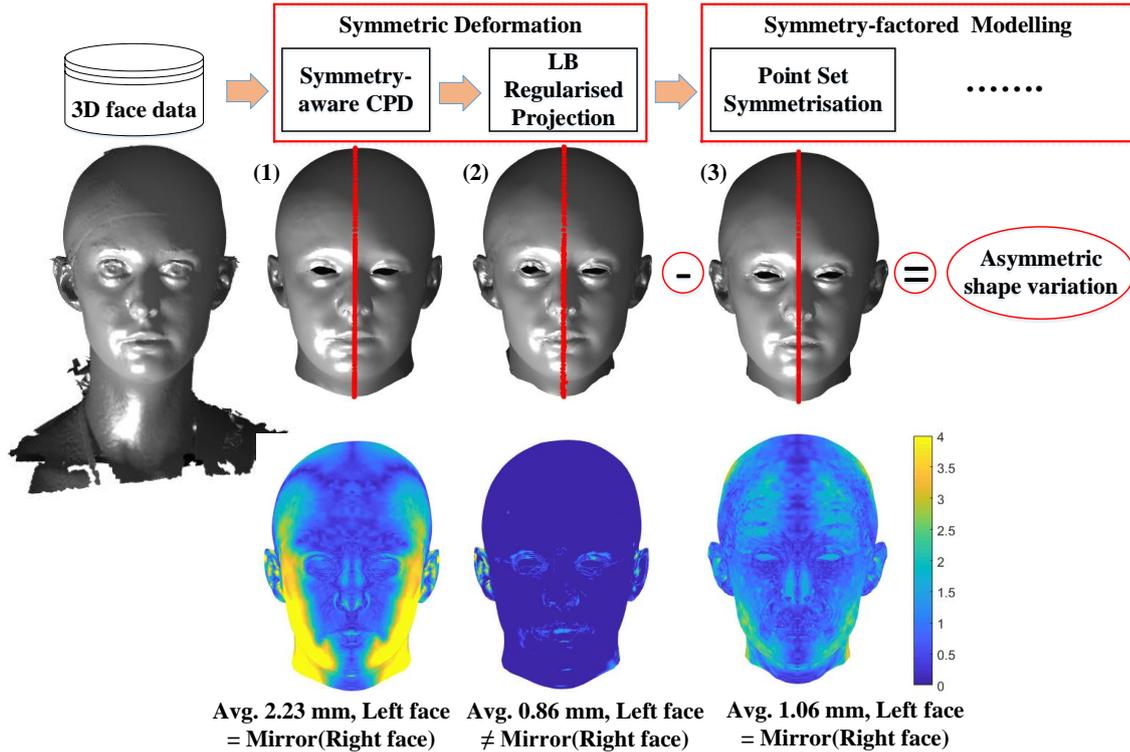


Figure 5.3: End results of symmetry-plane aware CPD, LB regularised projection and face symmetrisation. The color maps are the per-vertex nearest point error calculated from these results to target scan.

method in section 5.4. The constructed 3DMM of symmetry and asymmetry is evaluated in section 5.5. Finally Section 5.6 concludes the chapter.

## 5.1 Notation and Preliminaries

We represent a mesh either as a matrix of vertex positions  $\mathbf{X} \in \mathbb{R}^{N \times 3}$  or as a vector  $\mathbf{x} = \text{vec}(\mathbf{X}) = [x_1, y_1, z_1, \dots, z_N]^T$ , where  $\mathbf{x} \in \mathbb{R}^{3N}$ . The  $i$ th vertex in the mesh,  $\mathbf{p}_i \in \mathbb{R}^3$ ,  $i \in [1, N]$  is given by  $\mathbf{p}_i = [x_{3i-2}, x_{3i-1}, x_{3i}]^T$ . We assume that the meshes have been symmetrically remeshed. This means that for any vertex  $i$ , its symmetric partner is given by  $\text{sym}(i)$ . Vertices lying on the symmetry line are self-symmetric, i.e.  $i = \text{sym}(i)$ .

Therefore, each mesh is composed of  $N = 2M + S$  vertices,  $S$  of which are self-symmetric, leaving  $M = (N - S)/2$  pairs of proper symmetric vertices. Without loss of generality, we assume that the ordering of vertices is such that the proper-symmetric

vertices on one side of the mesh (e.g. left) come first, followed by the self-symmetric vertices and finally the proper-symmetric vertices on the other (e.g. right) side. Hence:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{\text{left}} \\ \mathbf{x}_{\text{self}} \\ \mathbf{x}_{\text{right}} \end{bmatrix},$$

with

$$\mathbf{x}_{\text{left}} = \mathbf{S}_{\text{left}}\mathbf{x}, \quad \mathbf{x}_{\text{self}} = \mathbf{S}_{\text{self}}\mathbf{x}, \quad \mathbf{x}_{\text{right}} = \mathbf{S}_{\text{right}}\mathbf{x},$$

where

$$\begin{aligned} \mathbf{S}_{\text{left}} &= \begin{bmatrix} \mathbf{I}_{3M} & \mathbf{0}_{3M \times 3(M+S)} \end{bmatrix} \in \{0, 1\}^{3M \times 3N}, \\ \mathbf{S}_{\text{right}} &= \begin{bmatrix} \mathbf{0}_{3M \times 3(M+S)} & \mathbf{I}_{3M} \end{bmatrix} \in \{0, 1\}^{3M \times 3N}, \\ \mathbf{S}_{\text{self}} &= \begin{bmatrix} \mathbf{0}_{3S \times 3M} & \mathbf{I}_{3S} & \mathbf{0}_{3S \times 3M} \end{bmatrix} \in \{0, 1\}^{3S \times 3N} \end{aligned}$$

are selection matrices that select the proper symmetric vertices from the left and right halves of the mesh and the self symmetric vertices respectively. The vectors  $\mathbf{x}_{\text{left}}$  and  $\mathbf{x}_{\text{right}}$  are assumed to appear in symmetry pair order and so the symmetry operator has a very simple form:

$$\text{sym}(i) = \begin{cases} i + M + S & \text{if } 1 \leq i \leq M \\ i & \text{if } M + 1 \leq i \leq M + S \\ i - M - S & \text{if } M + S + 1 \leq i \leq N \end{cases}$$

A symmetrised mesh is one with exact extrinsic symmetry. For clarity, when we refer to a mesh that has been symmetrised, we use a tilde,  $\tilde{\mathbf{x}}$ . We assume (without loss of generality) that such symmetrised meshes have their symmetry planes aligned with the  $x = 0$  plane. A single vertex is reflected about the  $x = 0$  plane by premultiplying with the reflection matrix:

$$\mathbf{F} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

so that  $\mathbf{F}\mathbf{v}$  is the reflection of  $\mathbf{v}$ . This can be extended to the whole template by

$$\mathbf{G}(M) = \mathbf{I}_M \otimes \mathbf{F}$$

so that  $\mathbf{G}(N)\mathbf{x}$  is the reflection of  $\mathbf{x}$  (we use  $\otimes$  to denote the Kronecker product).

To avoid redundancy in symmetrised meshes, we store only the vertices on one side of the mesh since those on the other can be reconstructed by reflection. Moreover, we need only store the  $y$  and  $z$  coordinates for the self-symmetric vertices since, by definition, their  $x$  coordinate is zero. Hence, if  $\tilde{\mathbf{x}}$  is a symmetrised mesh, we select the non-redundant entries by:

$$\mathbf{x}_{sym} = \begin{bmatrix} \mathbf{S}_{\text{left}} \\ \mathbf{S}_{\text{self},yz} \end{bmatrix} \tilde{\mathbf{x}} \in \mathbb{R}^{3M+2S}$$

where

$$\mathbf{S}_{\text{self},yz} = \begin{bmatrix} \mathbf{0}_{2S \times 3M} & \mathbf{I}_S \otimes \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \mathbf{0}_{2S \times 3M} \end{bmatrix}$$

is a selection matrix that selects only the  $y$  and  $z$  components of the self-symmetric vertices.

We can reconstruct a complete mesh from its reduced representation  $\mathbf{x}_{sym}$  using:

$$\tilde{\mathbf{x}} = \mathbf{T} \mathbf{x}_{sym} \quad (5.1)$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{I}_{3M+3S} \\ \mathbf{G}(M) \mathbf{S}_{-\text{self}} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3M} & \mathbf{0}_{3M \times 2S} \\ \mathbf{0}_{3S \times 3M} & \mathbf{I}_S \otimes \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix}$$

and

$$\mathbf{S}_{-\text{self}} = \begin{bmatrix} \mathbf{I}_{3M} & \mathbf{0}_{3M \times 3S} \end{bmatrix} \in \{0, 1\}^{3M \times 3(M+S)}$$

is a selection matrix that removes the self-symmetric vertices.

We can apply the same notation and principles to symmetric point motion, in addition to the point positions themselves, in order to facilitate symmetry maintenance during deformation. Thus we represent the template motion field as a matrix  $\mathbf{V} \in \mathbb{R}^{N \times 3}$ . or as a vector  $\mathbf{v} \in \mathbb{R}^{3N} = \text{vec}(\mathbf{V}) = [v_{x_1}, v_{y_1}, v_{z_1}, \dots, v_{z_M}]^T$ . The  $i$ th vertex motion in the template,  $\mathbf{v}_i \in \mathbb{R}^3, n \in [1, N]$  is given by  $\mathbf{v}_n = [v_{3n-2}, v_{3n-1}, v_{3n}]^T$ . Let any residual local template motions,  $\mathbf{v}$ , not explained by globally-symmetric deformations  $v_{sg}(\mathbf{Y})$ , be defined as;

$$\mathbf{v} = v(\mathbf{Y}) - v_{sg}(\mathbf{Y}) = \begin{bmatrix} \mathbf{v}_{\text{left}} \\ \mathbf{v}_{\text{self}} \\ \mathbf{v}_{\text{right}} \end{bmatrix} \quad (5.2)$$

and we can use the same selection matrices for vertex motions, as defined earlier for vertex positions.

$$\mathbf{v}_{\text{left}} = \mathbf{S}_{\text{left}}\mathbf{v}, \quad \mathbf{v}_{\text{self}} = \mathbf{S}_{\text{self}}\mathbf{v}, \quad \mathbf{v}_{\text{right}} = \mathbf{S}_{\text{right}}\mathbf{v} \quad (5.3)$$

## 5.2 Symmetric Shape Morphing

Our symmetric morphing work builds on Myronenko and Song [100], who derived the Coherent Point Drift (CPD) point registration algorithm in the context of both global affine deformations (CPD-affine) and local non-rigid deformations (CPD-nonrigid). They noted that non-rigid point set registration is an ill-posed problem and to obtain a unique solution, constraints on the solution space are required. They regularised the non-rigid motion field using a Gaussian kernel, so that deformation varies smoothly (‘drifts’ coherently) over the template surface. Our hypothesis is that a restriction to symmetric deformations may improve template morphing processes for (near) symmetric shapes; for example, it will not be possible for the sagittal symmetry contour of the template to deform via non-symmetric shearing and non-symmetric tangential surface sliding, which can occur in CPD-affine and CPD-nonrigid respectively.

If the shape that we are modelling has an *approximate* reflective symmetry, we can use this as a powerful constraint that can be leveraged in *any* morphable model construction pipeline. Thus our contribution is a new 3D shape template morphing algorithm that is suitable for any class of shapes that has regularity associated with a plane of reflective symmetry. The human face and full head are examples of such classes. Our approach requires strong but not perfect symmetry, as a final stage in our algorithm morphs the symmetrically deformed template to any asymmetries in the data. Therefore, our algorithm also permits the decomposition of shape into symmetric and asymmetric components, which is an interesting aspect of the study of shape variations and covariations within datasets. This is discussed in the following section. We call our symmetric deformation

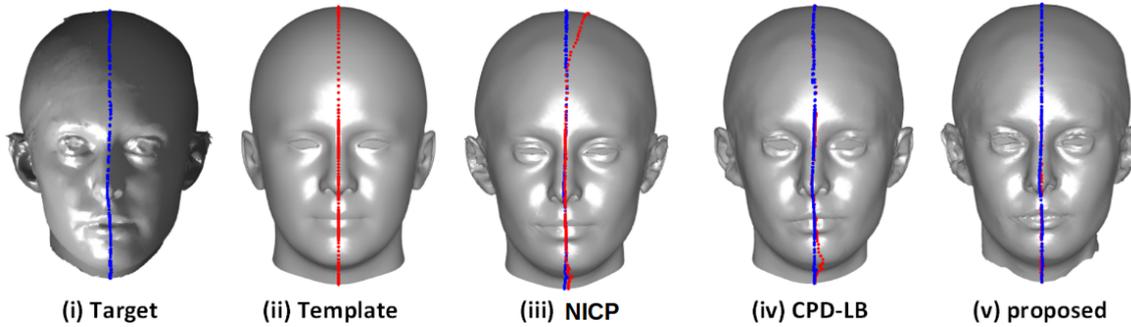


Figure 5.4: Symmetry contour comparisons after template deformation. Blue points are the pseudo ground truth symmetry contours on the target data (see section 5.3.1 for full explanation). Red points are from the symmetry contour indices of the source template. The deformed templates in (iii)-(v) show varying amounts of tangential sliding.

method *Symmetry-plane Aware CPD* (SA-CPD).

In SA-CPD, we symmetrise both CPD-affine and CPD-nonrigid. In the global affine case, we remove any non-symmetric shearing operations. In the local non-rigid case, we find the nearest symmetric deformation, in a least squares sense, to that generated by CPD-nonrigid. These symmetric deformations allow us to generate a symmetric template shape that is usually very close to the data. Finally, to account for any asymmetry in the data, we use a Laplace-Beltrami regularised projection operation, as presented in chapter 3.

Evaluations demonstrate that the proposed method outperforms many other template morphing (point registration) methods in the elimination of shape difference and sliding error. As can be seen in Figure 5.4, the shape difference in (iii)-(v) is small. However, (iii) has a large tangential sliding error, whereas (iv) and (v) have small sliding error, with (v) being the best. The proposed method can also deal with noise, outliers, and missing data. We also provide a means to perform gender and age classification from 3D shape. The proposed template deformation method gives the best performance in both gender and age classification tasks, as compared to other leading template deformation algorithms.

An overview of the process is shown in Figure 5.5. This consists of (i) a global symmetric deformation, which is a symmetrised affine transformation, derived from CPD-affine and (ii) a local symmetric deformation, derived from CPD-nonrigid. We can iterate the CPD-affine and CPD-nonrigid processes and this is indicated by the feedback path in

Figure 5.5. The motivation for this is that only relatively small non-rigid deformations are permitted between symmetrisations, and so the symmetry constraints acts as a kind of deformation regularisation. Note that small residual asymmetries can be accounted for by using a final regularised closest-point projection of the symmetrically deformed template onto the data mesh, which is the final step shown in Figure 5.5.

The remainder of this section describes the component processes in Figure 5.5 in more detail. In section 5.2.1, we describe the approximate input scan alignment. Then we present the global symmetric deformation in section 5.2.2 and local symmetric deformation in section 5.2.3. The LB regularised point projection is described in section 5.2.4.

### 5.2.1 Approximate Input Scan Alignment

The data in our Headspace dataset is not pose normalised and needs to be aligned to an approximate frontal pose, such that it approximately matches the pose of the template. This does not need to be accurate. It is sufficient that the initial alignment process reorientates the input scan such that it is within the convergence basin of CPD-affine. This alignment approach was described in the previous chapter and our paper [47]. In brief summary, our 3D input scans have an associated and registered colour-texture channel from which we detect 2D features using the approach of Zhu and Ramanan [157]. These 2D points are then projected to 3D points allowing pose normalisation by reorientating the detected 3D features to a template set of desired positions. The procedure was successful on all 1212 scans tested in the Headspace dataset.

### 5.2.2 CPD-affine for Global Symmetric Deformation

The global scale parameter in CPD’s so-called ‘rigid’ deformation formulation (it is actually a similarity transform) is often insufficiently general to give good deformation results. We prefer to use anisotropic scaling, which allows each dimension of the template to be scaled independently. Of course, the affine motion model can express this, but it also allows for XY and XZ shear, which are not permitted when the symmetry plane is at  $x = 0$ , although YZ shear is. The isotropically-scaled similarity motion model (termed ‘rigid’

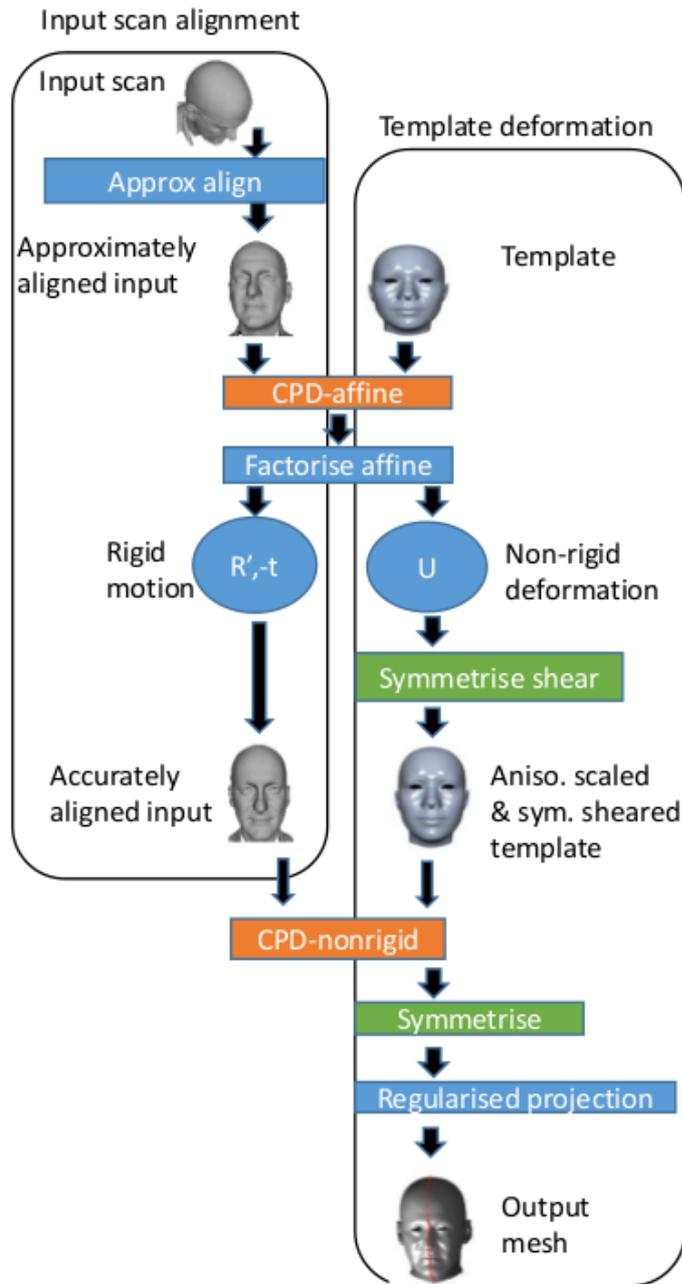


Figure 5.5: Symmetry aware CPD process. The left dashed outline shows alignment processes applied to the input data. The right dashed outline shows deformation processes applied to the template data. All but the final regularised projection are symmetric deformations.

in [100]) is given as:

$$T(\mathbf{y}_m; \mathbf{R}, \mathbf{t}, s) = s\mathbf{R}\mathbf{y}_m + \mathbf{t} \tag{5.4}$$

and the analysis to extract the optimal motion employs the orthogonality constraint  $\mathbf{R}^T \mathbf{R} = \mathbf{I}$ . If we augment the isotropic scale,  $s$ , to anisotropic scaling matrix  $\mathbf{S}_a = \mathbf{D}(s_x, s_y, s_z)$ , and we include a symmetric shear, we have a non-rigid symmetric transformation:

$$\mathbf{U}_{sx} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & m \\ 0 & 0 & s_z \end{bmatrix}, \quad (5.5)$$

then the motion model is more flexible than the similarity case, but is restricted to symmetric deformation and becomes:

$$T(\mathbf{y}_m; \mathbf{R}, \mathbf{U}_{sx}, \mathbf{t}) = \mathbf{R}\mathbf{U}_{sx}\mathbf{y}_m + \mathbf{t} \quad (5.6)$$

where the subscript  $sx$  denotes that the deformation is symmetric about the  $X = 0$  plane. We can optimise for CPD's global symmetric motion  $(\mathbf{R}, \mathbf{U}_{sx}, \mathbf{t})$  and variance  $(\sigma^2)$  parameters directly, but this is complicated as, in addition to the orthogonality constraint on  $\mathbf{R}$ , we need to handle the structural constraint on  $\mathbf{U}_{sx}$ . An alternative is to optimise with respect to a *general* affine motion, with the translation component expressed separately, i.e.  $\mathbf{T}_a = [\mathbf{B}, \mathbf{t}]$ . This is a more straightforward unconstrained optimisation, and the solution is presented in [100], which here we term CPD-affine. We then determine how to extract the nearest symmetric deformation to the general affine transformation. We achieve this by decomposing the affine transformation into a rigid part (a rotation) and a non-rigid part:

$$\mathbf{B} = \mathbf{R}\mathbf{U} \quad (5.7)$$

where  $\mathbf{U}$  is an upper-triangular matrix with anisotropic scalings on its diagonal and shears on its off-diagonal. Due to the orthogonality of  $\mathbf{R}$ , we have equivalent symmetric matrices such that

$$\mathbf{B}^T \mathbf{B} = \mathbf{U}^T \mathbf{U} \quad (5.8)$$

The known left side of the above equation is real and square-symmetric, and so we can form its Cholesky decomposition as:

$$\mathbf{B}^T \mathbf{B} = \mathbf{L}^T \mathbf{L} \quad (5.9)$$

and we set  $\mathbf{U} = \mathbf{L}^T$  as the upper-triangular matrix representing non-rigid deformation. We then extract the rotation matrix as

$$\mathbf{R} = \mathbf{B}\mathbf{U}^{-1} \quad (5.10)$$

Given we have  $\mathbf{U}$ , we can zero any non-symmetric shears in the X-Y and X-Z planes by zeroing the off-diagonal elements, in the first row of that matrix to give a deformation matrix  $\mathbf{U}_{sx}$ . Finally we can reconstruct the symmetrised affine matrix from its rigid and non-rigid parts as:

$$\mathbf{B} = \mathbf{R}\mathbf{U}_{sx} \quad (5.11)$$

Recalling the template deformation model from Equation 5.6, we split the update across the template and data such that the (inverse) rigid part of the affine update is applied to the data and the non-rigid part is applied to the template. The intent is to maintain the template in a frame where its sagittal symmetry plane is coincident with the  $X = 0$  plane, thus maintaining simplicity of form in any reflection matrix required for subsequent processing. Thus we update  $M$  template points  $y_m$  non-rigidly as:

$$\mathbf{y}_m \leftarrow \mathbf{U}_{sx}\mathbf{y}_m, \quad m = (1 \dots M) \quad (5.12)$$

and  $N$  data points  $x_n$  rigidly as:

$$\mathbf{x}_n \leftarrow \mathbf{R}^T(\mathbf{x}_n - \mathbf{t}), \quad n = (1 \dots N). \quad (5.13)$$

These operations are indicated in the process flow in Figure 5.5.

### 5.2.3 CPD-nonrigid for Local Symmetric Deformation

We now propose to find the nearest (in a least-squares sense) local symmetric deformation to the non-rigid component of any CPD-nonrigid deformation. For the required local shape deformation, we need to ensure that the displacement of proper symmetric point pairs is reflected across the symmetry plane. We assume the template maintains the pose of its symmetry plane on the  $YZ$  axis, with the inverse rigid motion being applied to the data, as described in the previous section.

Non-rigid CPD displaces the template,  $\mathbf{Y}$ , according to some displacement function,  $v$ :

$$\mathcal{T}(\mathbf{Y}, v) = \mathbf{Y} + v(\mathbf{Y}) \quad (5.14)$$

The general non-rigid motion can be considered to incorporate some (possibly zero) global-symmetric deformation. Hence, we decompose the displacement function as global

and local symmetric displacements:

$$v(\mathbf{Y}) \approx v_{sg}(\mathbf{Y}) + v_{sl}(\mathbf{Y}) \quad (5.15)$$

Our aim is to employ non-rigid CPD to generate the small displacements  $v(\mathbf{Y})$ , which can then be decomposed, to some approximation, into its symmetric global (sg) and symmetric local (sl) components. The vector field  $v(\mathbf{Y})$  will be a smooth motion field, with CPD-nonrigid using a Gaussian kernel to ensure smoothness. Any (small) global symmetric deformation,  $v_{sg}(\mathbf{Y})$ , contained within this can be determined from the process described in Sec. 5.2.2, using the template points before and after the non-rigid deformation as the initial points and target points of this incremental global-symmetric deformation respectively. Finally, we need to find an optimal, symmetrised, residual motion field  $v_{sl}(\mathbf{Y})$  in Equation 5.15, after  $v_{sg}(\mathbf{Y})$  is subtracted from  $v(\mathbf{Y})$ .

#### Proper Symmetric Deformation:

We define a reflection in the  $\mathbf{x} = 0$  symmetry plane by the matrix  $\mathbf{F}$ , where

$$\mathbf{F} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.16)$$

and we define  $\mathbf{v}_{left}^s$  as the *symmetric-left* non-rigid local motion field that we wish to recover (symmetric deformation is distinguished from non-symmetric by the superscript). The required symmetric-right motion is recovered by a reflection of this. The reflection can be applied to the  $P$  vertices on the left side of the template motion by

$$\mathbf{G}(P) = \mathbf{I}_P \otimes \mathbf{F} \quad (5.17)$$

so that

$$\mathbf{v}_{right}^s = \mathbf{G}(P)\mathbf{v}_{left}^s \quad (5.18)$$

is the reflection of  $\mathbf{v}_{left}^s$  (we use  $\otimes$  to denote the Kronecker product). Then we can formulate the computation of a proper symmetric motion field as:

$$\begin{bmatrix} \mathbf{I}_P \\ \mathbf{G}(P) \end{bmatrix} \mathbf{v}_{left}^s = \begin{bmatrix} \mathbf{S}_{left} \\ \mathbf{S}_{right} \end{bmatrix} \mathbf{v} \quad (5.19)$$

and we solve this linear LS problem for the symmetric-left motion  $\mathbf{v}_{left}^s$  and we recover the right symmetric motion as from Equation 5.18.

### Self-symmetric Deformation:

Finally, we require the motion of the self-symmetric points on the template symmetry plane to be restricted to that plane. The closest in-plane motion vectors to those of CPD-non-rigid are obtained by projecting to the  $x = 0$  plane with matrix,  $\mathbf{P}_x$

$$\mathbf{P}_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.20)$$

we define

$$\mathbf{P}_x(S) = \mathbf{I}_S \otimes \mathbf{P}_x \quad (5.21)$$

and the optimal  $S$  self-symmetric vertices  $\mathbf{v}_{sym_s}$  are computed as:

$$\mathbf{v}_{self}^s = \mathbf{P}_x(S) \mathbf{S}_{self} \mathbf{v} \quad (5.22)$$

## 5.2.4 Regularised Projection using Laplace-Beltrami

After symmetric template deformation, point projection to the aligned input data can eliminate any (normal) shape error. The template shape before and after this projection represents the symmetrised and non-symmetrised versions of template deformation respectively. Point projection is fragile if the input data is incomplete or noisy and may cause large artefacts. We overcome this by treating the projection operation as a mesh editing problem. We use the same method as described in Section 4.3.3. We write the LB mesh editing problem as a linear system of equations. Given the vertices of a data scan stored in the matrix  $\mathbf{X} \in \mathbb{R}^{N \times 3}$  and the deformed template obtained by CPD whose vertices are stored in the matrix  $\mathbf{Y} \in \mathbb{R}^{M \times 3}$ , we define the selection matrices  $\mathbf{S}_1 \in [0, 1]^{Q \times M}$  and  $\mathbf{S}_2 \in [0, 1]^{Q \times N}$  as those that select the  $Q$  vertices with mutual nearest neighbours from deformed template and data respectively. This linear system can be written as:

$$\begin{pmatrix} \lambda \mathbf{L} \\ \mathbf{S}_1 \end{pmatrix} \mathbf{Y}_{proj} = \begin{pmatrix} \lambda \mathbf{L} \mathbf{Y} \\ \mathbf{S}_2 \mathbf{X} \end{pmatrix} \quad (5.23)$$

where  $\mathbf{L} \in \mathbb{R}^{M \times M}$  is the cotangent Laplacian approximation to the LB operator [121] and  $\mathbf{Y}_{proj} \in \mathbb{R}^{M \times 3}$  are the projected vertex positions that we wish to solve for. The parameter  $\lambda$  weights the relative influence of the position and regularisation constraints, effectively

determining the ‘stiffness’ of the projection. As  $\lambda \rightarrow 0$ , the projection tends towards nearest neighbour projection. As  $\lambda \rightarrow \infty$ , the deformed template will only be allowed to rigidly transform.

## 5.3 Evaluation of Symmetric Deformation

In order to evaluate the symmetric deformation method described in section 5.2, we compare it with several other deformation methods in the literature. This is performed both qualitatively and quantitatively, using 1212 3D images in the Headspace dataset [47], which is publicly available. The following subsections describe: 1) qualitative and quantitative tangential sliding evaluation in section 5.3.1, 2) robustness to noise in section 5.3.2, and 3) gender and age classification performance using SVMs in section 5.3.3.

### 5.3.1 Tangential Sliding Evaluation

**Qualitative Evaluation:** We compare our SA-CPD method with NICP [8], the LSFM pipeline [27] (an NICP extension [8]), Li’s method [92], standard CPD (affine and non-rigid) [100] and CPD-LB [47]. Figure 5.6 shows a typical example where the proposed symmetric morphing method is qualitatively superior to other methods in terms of tangential sliding. All methods excluding the proposed and CPD-affine have observable tangential sliding problems. However, CPD-affine by itself significantly underfits to the target shape, some form of more flexible yet non-sliding deformation is required, as is provided by our method. We built a 3DMM using SA-CPD, as shown in Figure 5.7, the symmetry contour is stable in the middle when the shape is varied from +3SDs to -3SDs over the first ten principal components. This validates that the proposed method significantly mitigates tangential sliding over the full dataset.

**Quantitative Evaluation:** Pseudo ground truth symmetry contours are shown in blue in Figure 5.4 and can be compared to the template sagittal symmetry plane contour, shown in red. We compare our method with the LSFM pipeline [27] and CPD-LB [47]. Since the correspondence between the template and data target is unknown, it is not possible to compute the correspondence error directly. Instead, we employ two metrics: 1) the Nearest Point Error (NPE) to quantify the shape difference from the deformed template to

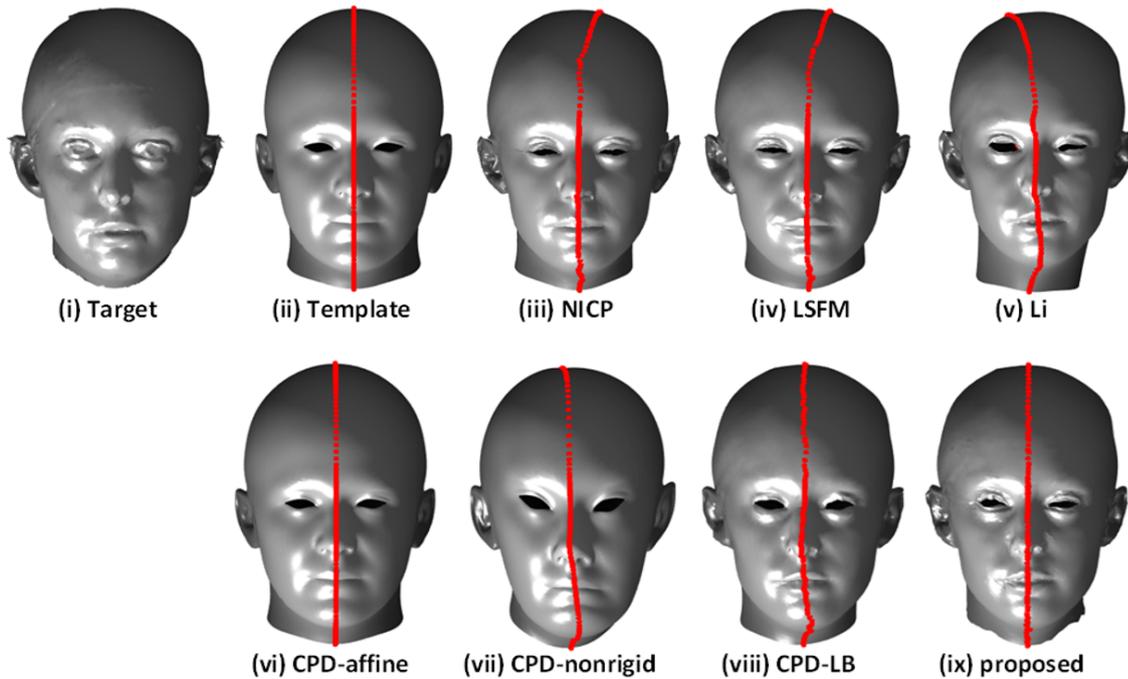


Figure 5.6: Deformation of the template to the first scan using competing methods. Note the tangential sliding in all methods except the proposed and CPD-affine. CPD-affine is likely to have some small shear and significantly underfits, but the proposed method has an excellent fit to the target data.

the target; 2) Symmetry Contour Error (SCE) to quantify the tangential sliding error. The NPE is computed by measuring the nearest point distance from the deformed template to raw scan and averaging over all vertices. As illustrated in Figure 5.8 (a), 87% of the NPE from our method is under 1mm, which compares to 30% for CPD-LB and 28% for the LSFM pipeline. We use piecewise-trimmed ICP between the raw scan and its reflection [106] to detect the local symmetry contour (blue contour in Figure 5.4) in the raw scan and we use this as a pseudo ground truth. This allows us to compute the SCE metric. (This blue symmetry contour is far less subject to surface sliding problems as it employs local-piecewise registration of the data to its self-reflection [16], and it employs robust outlier rejection. This contour can track local asymmetries, such as the nose bending to the left/right.) Figure 5.8 (b) shows that 99% of SCE from our method is under 2mm, which compares to 82% for CPD-LB and 0.6% for LSFM. Overall, the proposed method significantly outperforms the other two methods across both metrics.

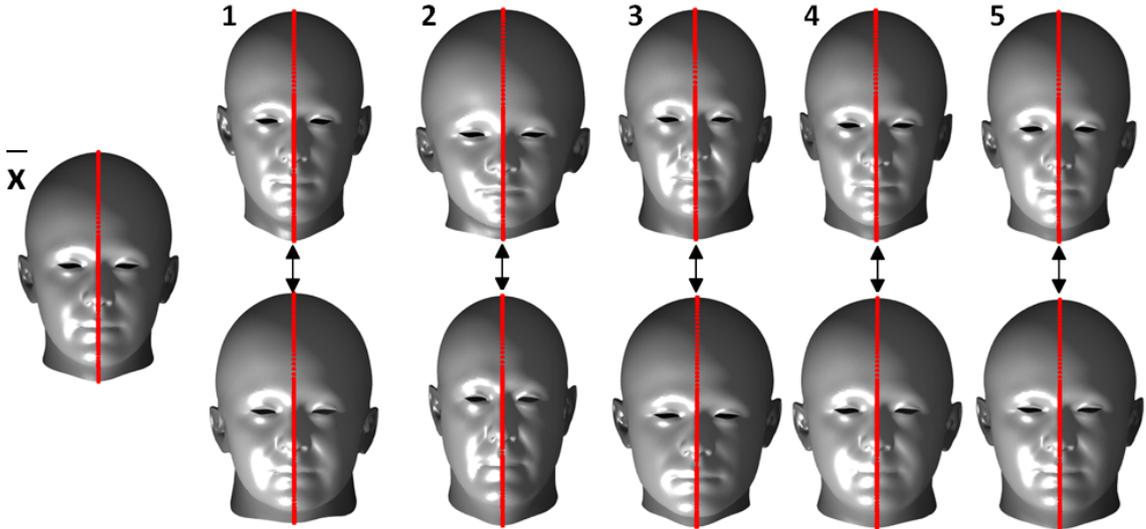


Figure 5.7: 3D morphable model constructed by SA-CPD. The mean and the first five principal components are shown at +3SD (top row) and -3SD (bottom row). Note the stability of the symmetry contour, with no tangential sliding across the main eigenvectors.

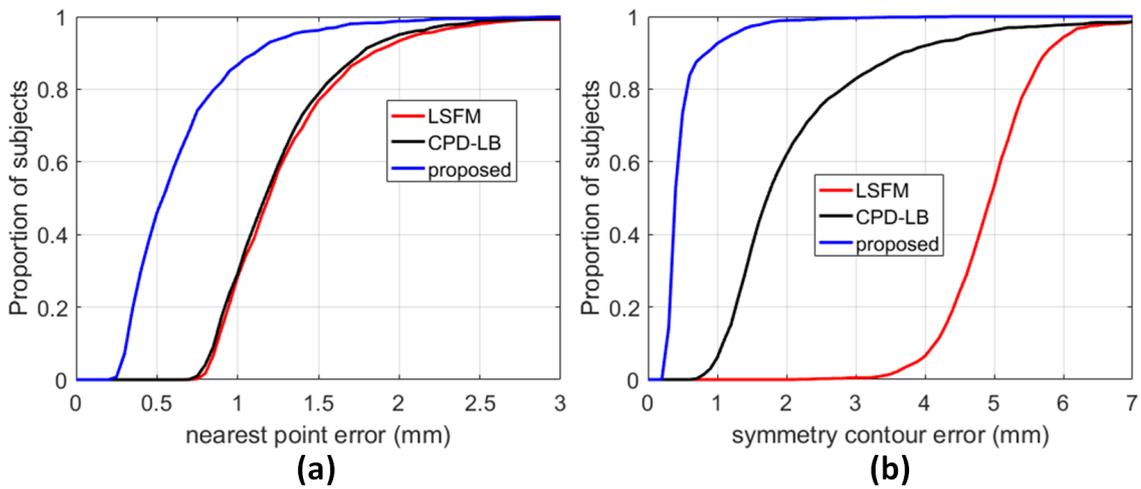


Figure 5.8: (a) Proportion of subjects with a Nearest Point Error (NPE) less than abscissa value. (b) Proportion of subjects with a Symmetry Contour Error (SCE) less than abscissa value.

### 5.3.2 Robustness

We use a 3D data mesh with outliers, missing data and Gaussian noise to test the robustness of the proposed method. When dealing with these situations, deformation algorithms need to choose the proper parameters. So in this section, it is unfair to compare other algo-

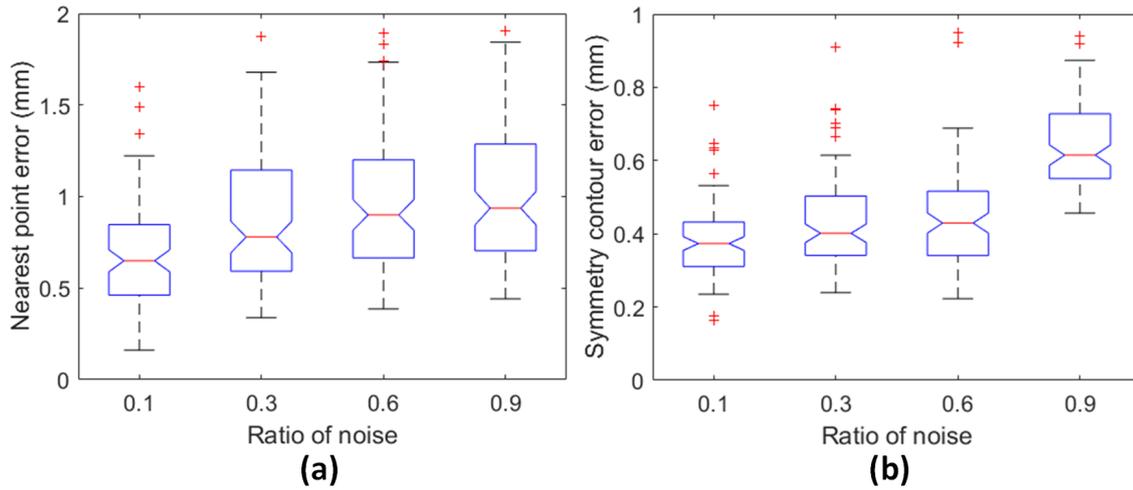


Figure 5.9: (a) NPE and (b) SCE for 100 3D data scans against level of Gaussian noise.

Table 5.1: Gender classification results

	Precision	Recall	F-score
LSFM	0.79	0.80	0.79
CPD-LB	0.81	0.81	0.81
Proposed	0.84	0.84	0.84

rithms with the proposed method, without extensive parameter tuning. We add Gaussian noise data to 100 3D meshes in the dataset. The mean of the Gaussian noise is set at the mean of the target data and variance is set to be compatible with head size, by scaling a unit normal distribution by 80mm, as shown by the blue points in Figure 5.10, row (3). We define ‘ratio of noise’ as the number of Gaussian noise points as a fraction of the number of template points,  $M$ . In Figure 5.9, we demonstrate the NPE of the proposed method when dealing with different percentages of Gaussian noise. Figure 5.10 shows the qualitative results of the proposed method when dealing with outliers, missing data and Gaussian noise (ratio 0.6) along with error metric computations. Overall, the proposed SA-CPD method is shown to be robust to outliers, missing data and Gaussian noise.

### 5.3.3 Gender and Age Classification

We use the deformation results of the proposed SA-CPD method, LSFM pipeline [27], and CPD-LB [47] to build three 3D morphable models. Then all of the 1212 face meshes

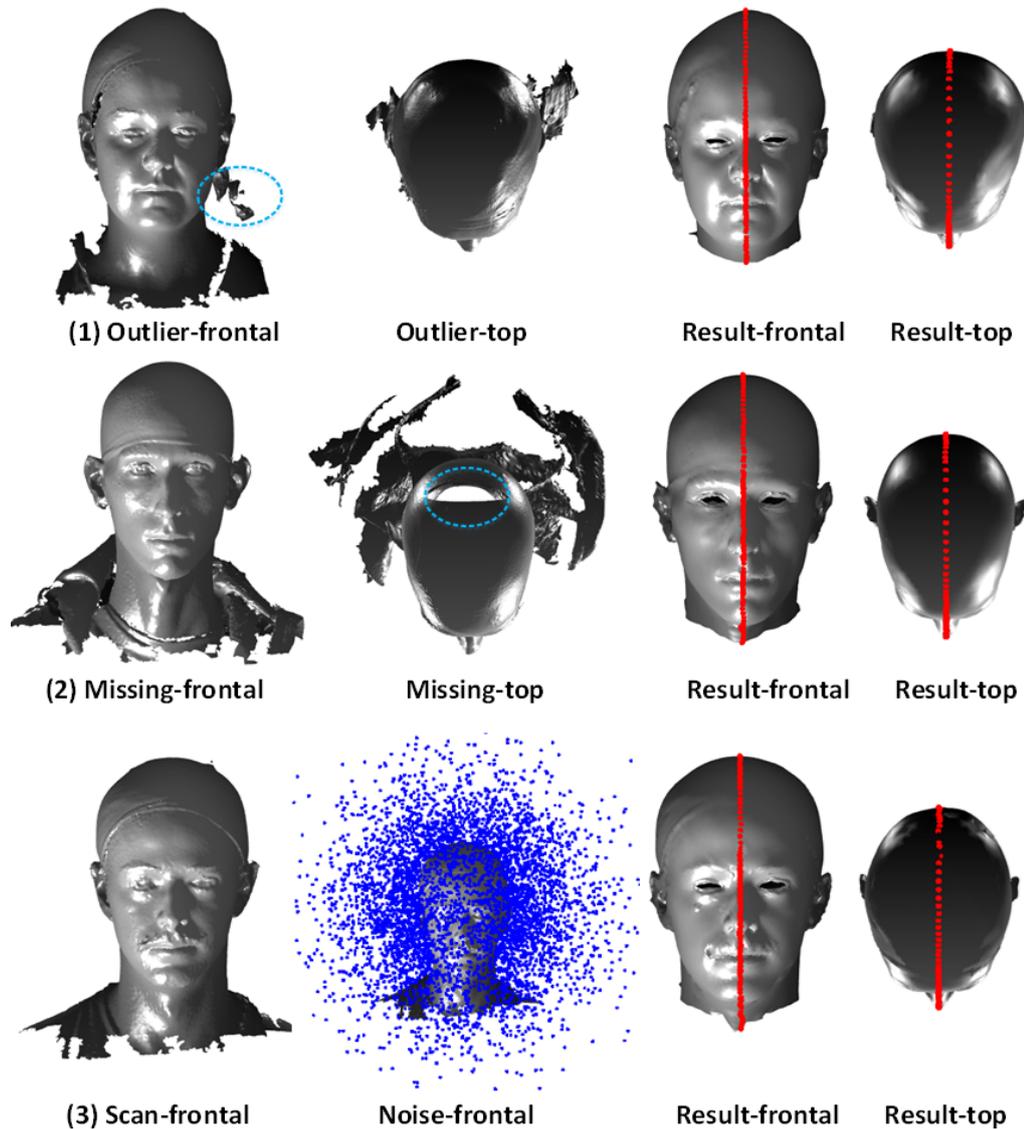


Figure 5.10: Deformation results against: (1) outlier, NPE = 1.3023, SCE = 0.2843; (2) cranial data missing, NPE = 0.4418, SCE = 0.3081; (3) Gaussian noise (ratio 0.6), NPE = 0.9342, SCE = 0.6992.

in the dataset are reparameterised using each of the the models. Using the demographic information (metadata) within the Headspace dataset we train a Support Vector Machine (SVM) classifier for each model, which maps the corresponding shape vectors to the gender groups and four age groups (0-11, 12-21, 22-60 and over 60). To measure the classification accuracy, we use the classifier to predict the age bracket and the gender for the test subjects via a 10-fold cross-validation evaluation so that no test subject ever appears in the classifier’s training set. This provides an application-oriented evaluation of

Table 5.2: Age classification results

	Precision	Recall	F-score
LSFM	0.71	0.71	0.71
CPD-LB	0.72	0.72	0.72
Proposed	0.73	0.74	0.73

the quality of the correspondence and low-dimensional representation. As can be seen in Table. 5.1 and 5.2, the proposed SA-CPD deformation method has the best performance in both gender and age classification.

## 5.4 Symmetry-factored Statistical Modelling of Cranio-facial Shape

Here we present the first 3D Morphable Model of craniofacial symmetry and asymmetry. In our proposed symmetry-factored statistical modelling pipeline, we employ five processes to model the symmetric and asymmetric deformations as follows:

**Section 5.4.1 Symmetry Detection and Symmetric Remeshing:** Given an arbitrary mesh, for every vertex attempt to find a corresponding symmetric partner vertex or point on the surface. Then remesh the surface with symmetric topology so that every vertex has a symmetric partner or is self-symmetric. This is done by the proposed symmetric deformation method described in the previous section.

**Section 5.4.2 Symmetry Plane Alignment:** Given a mesh with symmetric topology, estimate a plane which maximises the symmetry about the plane.

**Section 5.4.3 Point Set Symmetrisation:** Given a mesh with symmetric topology, find the ‘closest’ shape with exact geometric symmetry.

**Section 5.4.4 Symmetry-plane Aware GPA:** Given a collection of meshes with symmetric topology with symmetry planes aligned, solve for the remaining degrees of freedom to best align the collection.

**Section 5.4.5 Symmetry-factored Statistical Modelling:** Separately model symmetric and asymmetric modes of shape variation.

Our specific aim in this section is to model deviations from exact extrinsic symmetry for a set of 3D head scans with neutral facial expression.

### 5.4.1 Symmetry Detection and Remeshing

Symmetry detection and remeshing can be done simultaneously using template morphing methods and we apply the symmetric deformation method in section 5.2. This process is able to avoid over-fitting and under-fitting and these two properties ensure that the symmetry detection and remeshing is reliable.

### 5.4.2 Symmetry Plane Alignment

For a given mesh, we wish to find a best-fit symmetry plane that maximises bilateral extrinsic symmetry. Denote a plane with equation:

$$ax + by + cz = d \quad (5.24)$$

by  $\mathbf{p} = [a, b, c, d]$ . We denote by  $\mathbf{p} : \mathbb{R}^3 \mapsto \mathbb{R}^3$  the reflection of a point about the plane given by  $\mathbf{p}$ . We seek the plane which satisfies the following optimisation problem:

$$\operatorname{argmin}_{\mathbf{p}} \sum_i \|\mathbf{p}(\mathbf{vert}_i) - \mathbf{vert}_{\mathit{sym}(i)}\|^2. \quad (5.25)$$

This problem can be solved in closed form in a straightforward manner. First we compute a modified point cloud in which we take the average of each vertex and its symmetric partner. This gives an almost planar point cloud (it would be exactly planar for a mesh with exact extrinsic symmetry). Second, we apply PCA to this modified point cloud and select the eigenvector with the smallest eigenvalue to give the plane normal. The centre of mass gives a point on the plane. Finally, having found the symmetry plane  $\mathbf{p}$ , compute a rigid transformation  $[\mathbf{R}, \mathbf{t}]$  such that the symmetry plane coincides with the  $x = 0$  plane.

### 5.4.3 Point Set Symmetrisation

We pose symmetrisation as finding the symmetric mesh that minimises distortion relative to a given mesh,  $\mathbf{x}$ , where distortion is measured via linear operators  $\mathbf{M}_a, \mathbf{M}_b \in R^{3N \times 3N}$ .



and an optional scale  $s$ . To factor out translation, all samples (and the mean) have their  $y, z$  centre of mass subtracted:

$$t_y = -\frac{1}{N} \sum_{i=1}^N \mathbf{X}_{2,i}, \quad (5.29)$$

$$t_z = -\frac{1}{N} \sum_{i=1}^N \mathbf{X}_{3,i}, \quad (5.30)$$

Ignoring scale for now and assuming that  $\mathbf{X}$  and  $\bar{\mathbf{X}}$  have had their  $y, z$  centre of mass subtracted, we solve the following optimisation problem in terms of the angle of rotation:

$$\begin{aligned} & \arg \min_{\theta} \|\mathbf{R}_x(\theta)\mathbf{X} - \bar{\mathbf{X}}\|_{\text{Fro}}^2 \\ &= \arg \min_{\theta} \sum_{i=1}^N (\cos \theta \mathbf{X}_{2,i} - \sin \theta \mathbf{X}_{3,i} - \bar{\mathbf{X}}_{2,i})^2 \\ & \quad + (\sin \theta \mathbf{X}_{2,i} + \cos \theta \mathbf{X}_{3,i} - \bar{\mathbf{X}}_{3,i})^2 \end{aligned}$$

Differentiating with respect to  $\theta$ , setting to zero and solving for  $\theta$  yields:

$$\begin{aligned} \mathbf{F}_1 &= \sum_{i=1}^N (\mathbf{X}_{2,i} \bar{\mathbf{X}}_{3,i} - \mathbf{X}_{3,i} \bar{\mathbf{X}}_{2,i}), \\ \mathbf{F}_2 &= \sum_{i=1}^N (\mathbf{X}_{2,i} \bar{\mathbf{X}}_{2,i} + \mathbf{X}_{3,i} \bar{\mathbf{X}}_{3,i}), \\ \theta &= \text{atan2}(\mathbf{F}_1, \mathbf{F}_2). \end{aligned}$$

Having aligned all of the meshes to the mean using the above steps, we compute a new mean and iterate. We refer to this process as Symmetry-plane aware GPA (SA-GPA).

### 5.4.5 Building a Symmetry-factored Model

We now build a statistical model composed of two parts: a symmetric part (which only need model points on one side of the symmetry plane, and the self-symmetric points on the symmetry plane, since the other half is by definition given by a reflection), and an asymmetric part (modelled as displacements to the output of the symmetric model).

We now construct a symmetric data matrix  $\mathbf{D}_{sym} \in \mathbb{R}^{(3M+2S) \times K}$  from the  $K$  symmetrised and SA-GPA aligned training meshes. We apply PCA to find eigenvectors/eigenvalues of the symmetrised meshes. The symmetric statistical model is given by:

$$\mathbf{x}_{sym} = \bar{\mathbf{x}}_{sym} + \mathbf{P}_{sym} \mathbf{b}_{sym} \quad (5.31)$$

where  $\mathbf{P}_{sym} \in \mathbb{R}^{(3M+2S) \times K}$ ,  $\bar{\mathbf{x}}_{sym} \in \mathbb{R}^{3M+2S}$ . A complete symmetric mesh can be built from  $\mathbf{x}_{sym}$  using (5.1).

The data matrix for the asymmetric model,  $\mathbf{D}_{asym} \in \mathbb{R}^{3N \times K}$ , is constructed by subtracting the symmetrised meshes from the original meshes, such that the  $i$ th column is given by:

$$\mathbf{D}_{asym,i} = \mathbf{x}_i - \mathbf{T}\mathbf{D}_{sym,i}. \quad (5.32)$$

We again apply PCA to this matrix but without computing and subtracting a mean since the data matrix directly describes displacements. The final combined model is hence given by:

$$\mathbf{x} = \mathbf{T}(\bar{\mathbf{x}}_{sym} + \mathbf{P}_{sym}\mathbf{b}_{sym}) + \mathbf{P}_{asym}\mathbf{b}_{asym} \quad (5.33)$$

Note that this could be rewritten in the form:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P} \begin{bmatrix} \mathbf{b}_{sym} \\ \mathbf{b}_{asym} \end{bmatrix} \quad (5.34)$$

where

$$\bar{\mathbf{x}} = \mathbf{T}\bar{\mathbf{x}}_{sym} \quad (5.35)$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{sym} & \mathbf{P}_{asym} \\ \mathbf{G}(M)\mathbf{S}_{-self}\mathbf{P}_{sym} & \end{bmatrix}. \quad (5.36)$$

Hence, it is still just a standard linear model but for which the parameters can be partitioned into those that only vary the shape symmetrically and those that vary it asymmetrically.

## 5.5 Evaluation of Symmetry-factored Modelling

We use the registered data from the proposed symmetric deformation algorithm to derive our symmetry-factored 3D craniofacial model, by applying the model construction pipeline, detailed in section 5.4. We compare the proposed model with linear PCA models which are symmetry-plane aware GPA + PCA and standard GPA + PCA. In section 5.5.1, we demonstrate our 3DMM of craniofacial symmetry and asymmetry. In section 5.5.2, we show the outcome of each step in symmetry-factored modelling and the model fitting error. The evaluation of the proposed 3DMMs is included in section 5.5.3. We

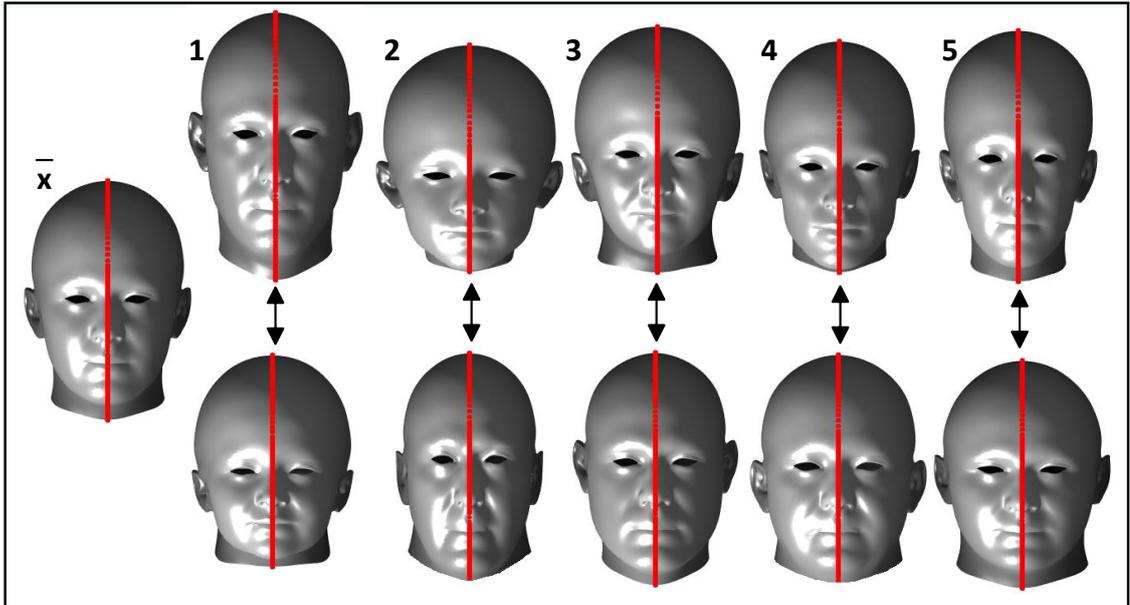


Figure 5.11: Visualisation of 5 principal components (mean  $\pm 5$  SDs) of symmetry variation.

compare linear averaging symmetrisation method with the proposed method in section 5.5.4. Section 5.5.5 describes the reconstruction from a half head mesh using the proposed 3DMMs. In section 5.5.6, we present age classification in terms of craniofacial symmetry and asymmetry.

### 5.5.1 3D Morphable Model of Asymmetry

Figure 5.11 shows the morphable model of symmetrical variation; specifically, mean  $\pm 5$  standard deviations of symmetrical variation; Figure 5.12 shows the morphable model of asymmetry variation: Top - without symmetric morphing; Bottom - with symmetric morphing (i.e. using SA-CPD). In order to observe pure asymmetry variation, we add in the mean of the symmetry model  $\pm 5$  standard deviations of asymmetrical variation. The red points are the symmetry contour. Note that the major mode of asymmetry is the angle between the head and neck. When not using symmetric morphing (SA-CPD), we can observe an obvious tangential sliding of the symmetry contour in the third mode of the morphable model. So there is tangential sliding motion in the asymmetric variation, which is not desirable for asymmetry modelling. With symmetric deformation, the symmetry

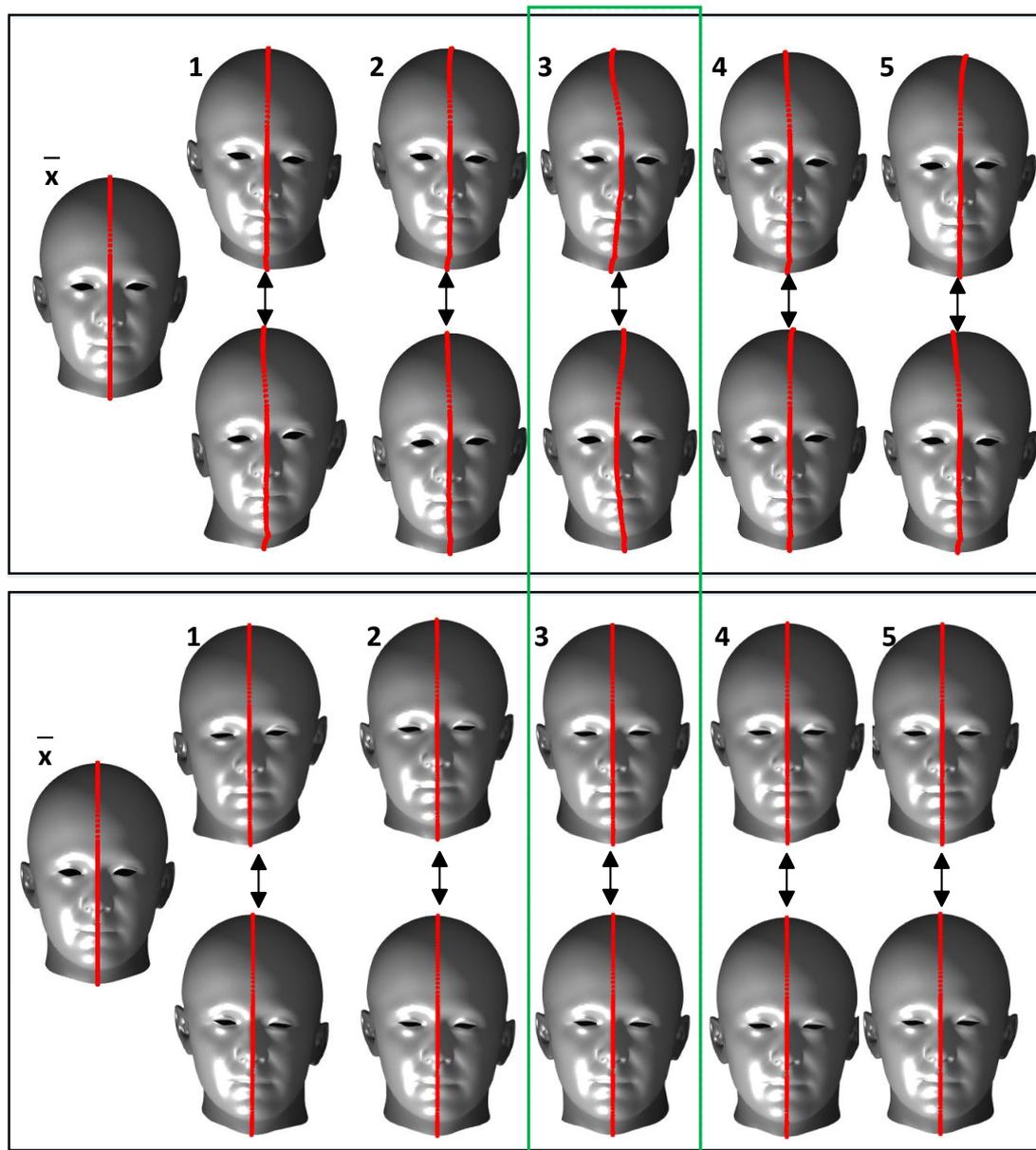


Figure 5.12: Visualisation of 5 principal components (mean  $\pm 5$  SDs) of asymmetry variation: Top - without symmetric deformation; Bottom - with symmetric deformation.

contour is fixed to the  $X = 0$  plane. So there is no asymmetric tangential sliding motion included in the proposed framework. This also validates that symmetric deformation is necessary for symmetry-factor modelling.

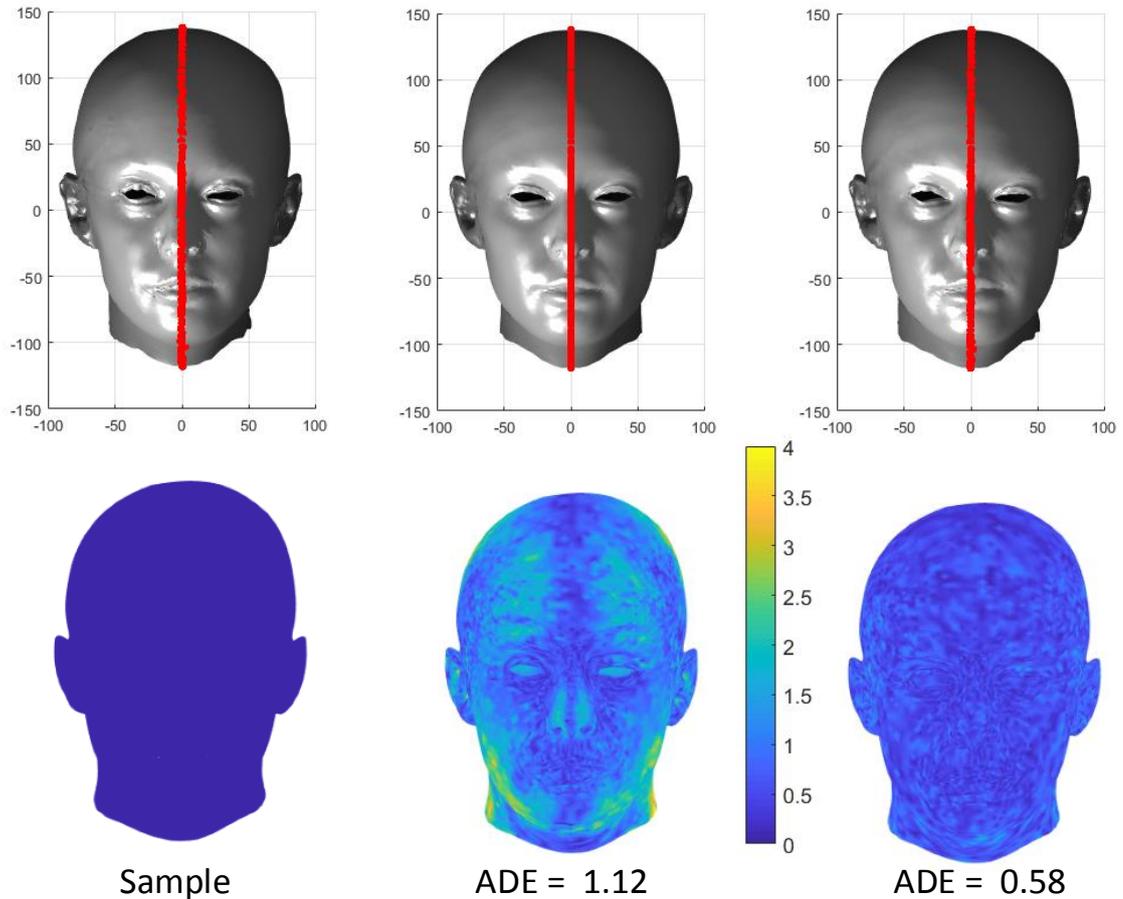


Figure 5.13: (1) Symmetry plane alignment, red line shows deformed symmetry contour from initial template morphing (2) model fitting using 3DMM of symmetry; (3) model fitting using 3DMM of symmetry + asymmetry; Average distance error (ADE, mm) against the shape sample (1).

### 5.5.2 Model Fitting

Figure 5.13 illustrates the end results of symmetry plane alignment and model fitting. Figure 5.13 (2) shows the reconstruction of symmetric face for this specific sample. The distance error color map becomes symmetrical, see Figure 5.13 (2). The symmetry plane of Figure 5.13 (2) is the  $X = 0$  plane. The aim of model fitting is to find the parameters  $\mathbf{P}$  in Equation 5.34 and validate the effectiveness of morphable models. Figure 5.14 (1) shows the cumulative error distributions of the per-vertex fitting error. For very small fitting errors (less than 0.487 mm), the standard PCA model is better than the proposed model, but for larger fitting errors, which is 91.3% of the dataset, the proposed model is

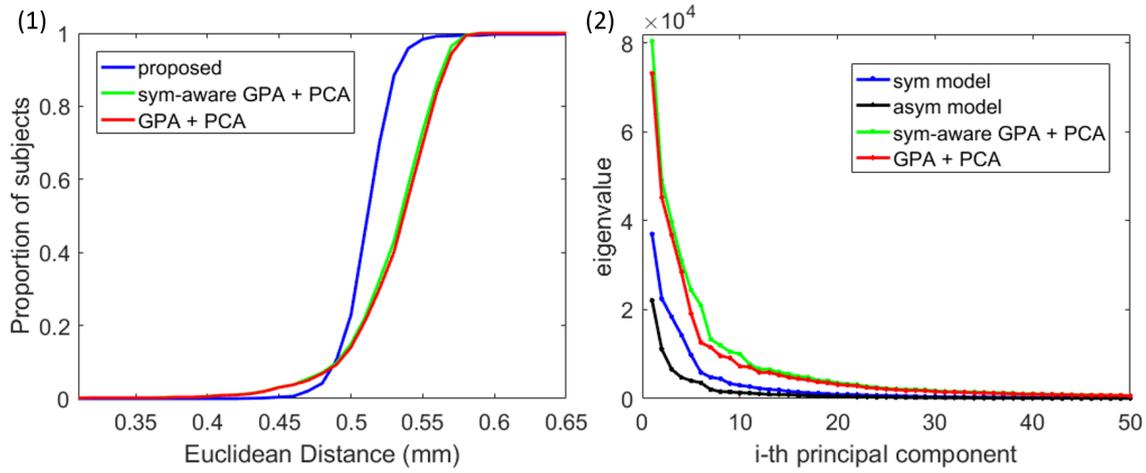


Figure 5.14: (1) Cumulative error distributions of the per-vertex fitting error (mm): proposed vs. standard PCA model (higher is better); (2) Plot of eigenvalues. Fewer PCs describe more variation.

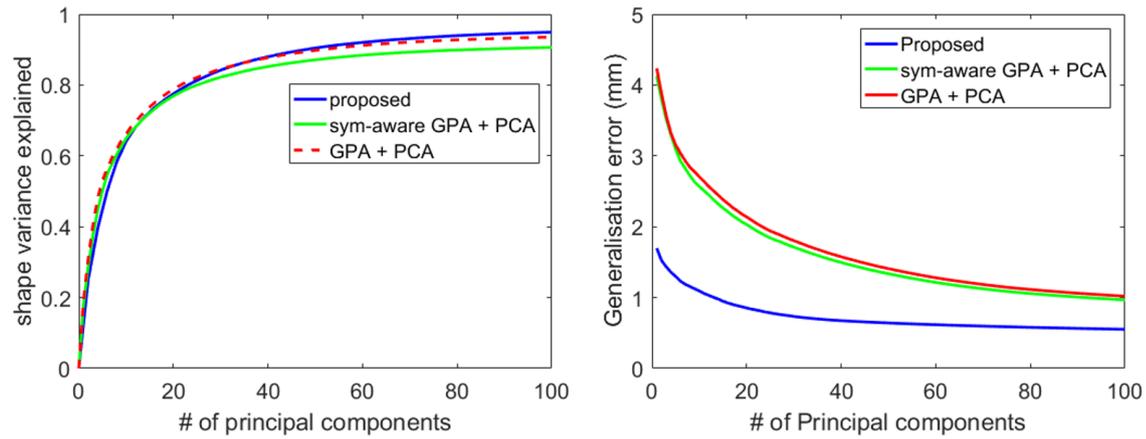


Figure 5.15: Left - compactness, higher is better; Right - generalisation error of the proposed model and PCA model with the number of principal components retained, lower is better.

better than the standard PCA models. The proposed symmetry-plane aware GPA improves the performance of the standard PCA model.

### 5.5.3 Model Evaluation

We compute model compactness and generalisation error in order to evaluate the proposed model quantitatively. To ensure a fair comparison, when we compute these two

criteria, the number of principal components (PCs) from the standard PCA model is the sum of PCs from symmetry-factored model, i.e. if we use  $N_{pc}$  PCs from the standard PCA model, the number of PCs from the symmetry model is  $\frac{N_{pc}}{2}$  and also  $\frac{N_{pc}}{2}$  from the asymmetry model. As can be seen from Figure 5.14 (2), the asymmetry model requires fewer components to express its variation than the symmetry model, which in turn requires fewer components than the PCA model. When we combine  $\frac{N_{pc}}{2}$  symmetry and  $\frac{N_{pc}}{2}$  asymmetry components with  $N_{pc}$  PCA components in the compactness plot in Figure 5.15, we find that the proposed model is slightly better than the standard PCA model and more compact than the symmetry-plane aware GPA + PCA model.

The generalization ability of a model measures its capability to represent unseen instances of the object class [123]. The generalization ability of each model is measured using leave-one-out reconstruction. A model is built using all but one member of the training set and then fitted to the excluded example. The generalisation error shown on the right of Figure 5.15 implies that the proposed model has significantly better generalisation ability than the standard PCA models. The symmetry-plane aware GPA + PCA model has slightly lower reconstruction error than the standard PCA model, which implies that the proposed symmetry-plane aware GPA can improve model generalization ability.

#### 5.5.4 Symmetrisation Comparison

Figure 5.16 (1) shows a subject with a slightly bending nose. Figure 5.16 (2) demonstrate linear averaging of the left and right halves of the mesh, and Figure 5.16 (3) demonstrate the proposed point set symmetrisation. The nose bridge becomes thinner with linear averaging. We also calculate the geodesic distance error for the symmetry pairs shown in Figure 5.16. Since the sample data with specific distortion is limited, we use 5 subjects with bending noses in Headspace dataset. When compared with symmetry pairs in sample data, the average Geodesic distance error for linear averaging is 0.42 cm and that for the proposed method is 0.19 cm.

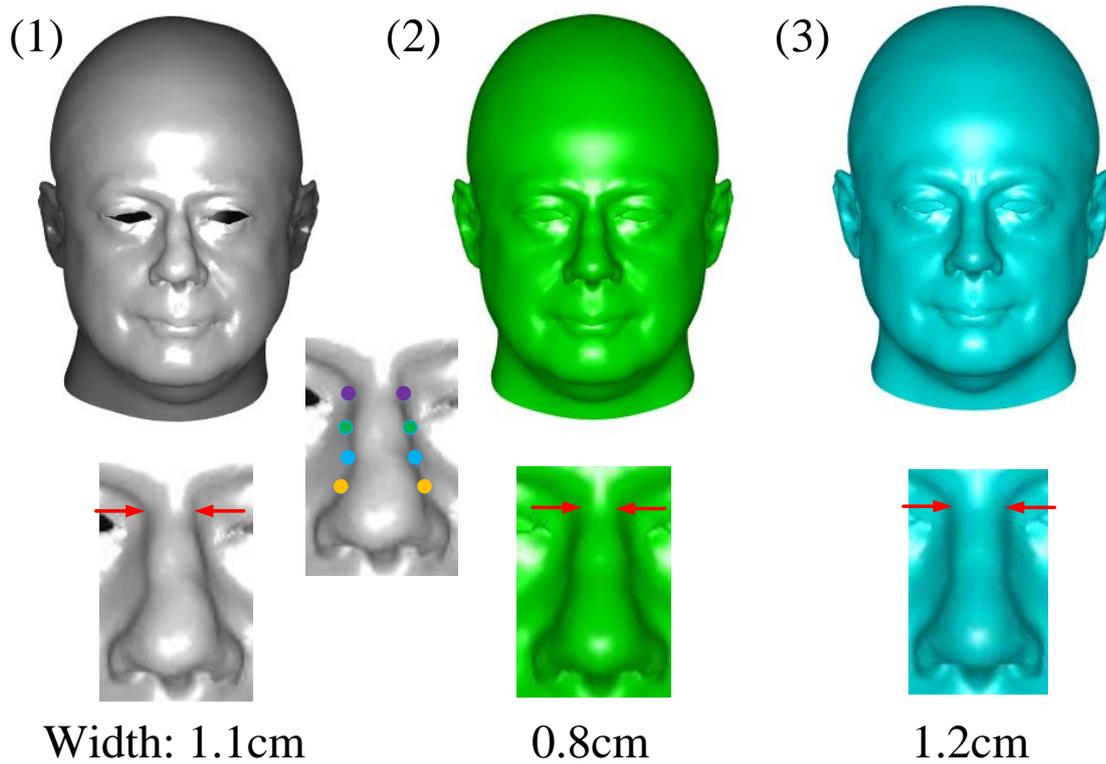


Figure 5.16: point set symmetrisation comparison: (1) Sample mesh with bending nose; (2) Linear average reduces width; (3) Proposed symmetrisation method has a smaller error.

### 5.5.5 Half Head Completion

When shade-from-shading is applied to a 2D profile image, the standard way to make the depth information complete is to calculate the reflection of the half head. There is no asymmetry variation in this completion. A PCA model can retain asymmetry variation in completion, but it tends to lose more symmetry information. The proposed model is able to overcome the loss of both symmetry and asymmetry information. From Equation 5.34, the reconstruction from the proposed model can be decomposed into two steps:

- (i) Find the closest symmetry head, computing the symmetry parameters  $\mathbf{P}_{sym}$  by least squares;
- (ii) Use  $\mathbf{P}_{sym}$  to find the closest asymmetry variation, computing the asymmetry parameters  $\mathbf{P}_{asym}$  by least squares.

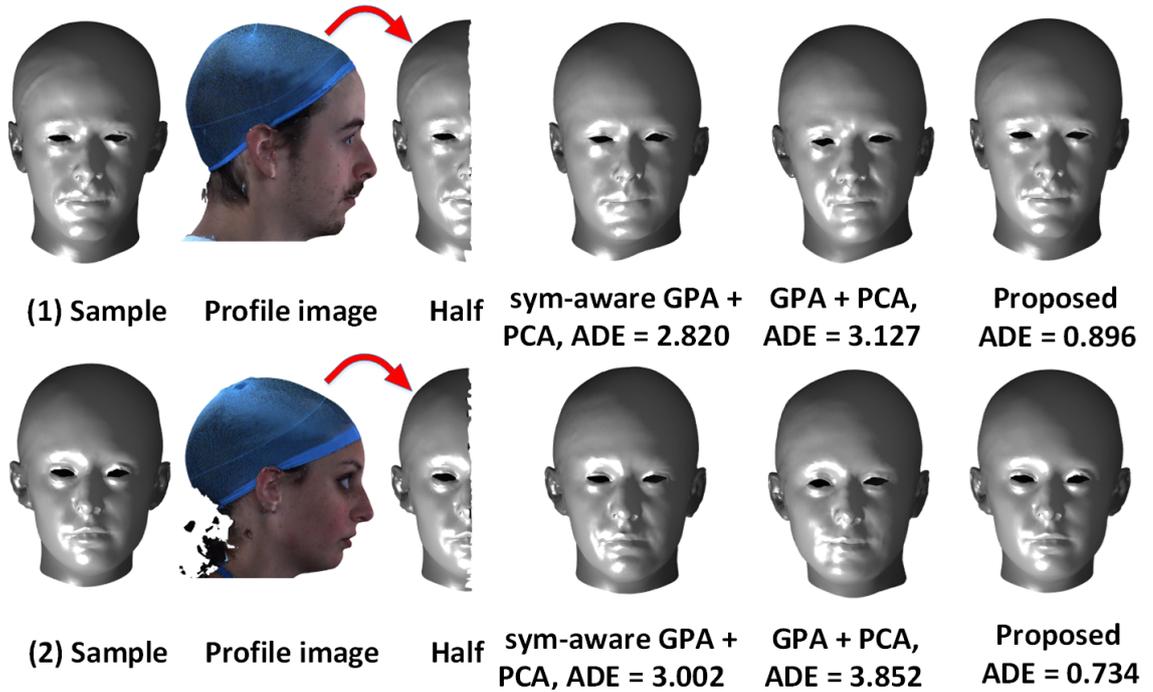


Figure 5.17: Simulation of half head completion from profile image.

Table 5.3: Age classification using shape parameters

	Precision	Recall	F-score
PCA	0.733	0.735	0.731
Sym	0.737	0.741	0.736
Asym	0.709	0.712	0.710
Sym+Asym	<b>0.739</b>	<b>0.741</b>	<b>0.741</b>

Step (i) retains symmetry information and step (ii) retains asymmetry information. This can be validated by the results in Figure 5.17 and Figure 5.18 in half-head completion experiments. We use 606 training samples to build the 3DMMs of craniofacial asymmetry and use the rest of dataset (606 subjects) for the half-head completion task. 200 PCs from PCA model are used to do the half head completion. For a fair comparison, 100 PCs from symmetry model and 100 PCs from asymmetry model are used. As shown in Figure 5.18, the proposed model has much greater proportion of subjects with lower reconstruction error from half head.

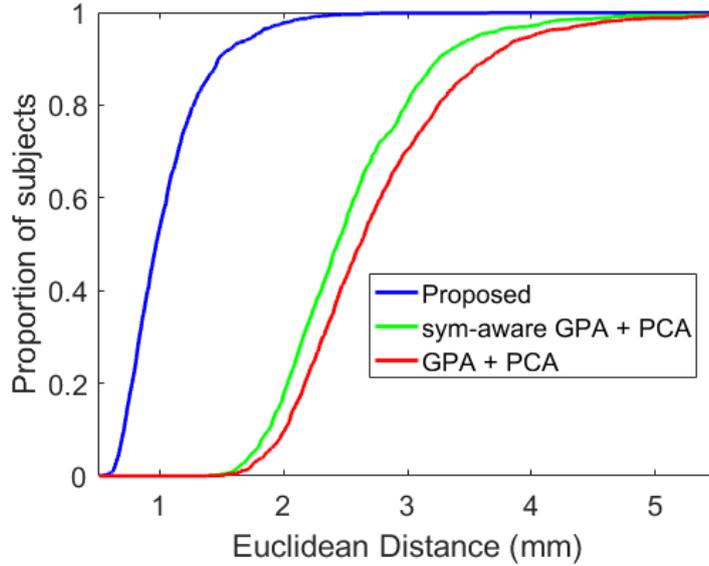


Figure 5.18: Cumulative error distributions of the per-vertex fitting error (mm) from half head: proposed model vs. standard PCA model (higher is better).

### 5.5.6 Age Classification: Asymmetry vs. Symmetry

We divide the dataset into four age groups: age 0-11, 12-21, 22-60 and  $> 60$ .  $\mathbf{P}_{sym}$  and  $\mathbf{P}_{asym}$  are used for the age classification task and compared with the standard PCA model. The comparison uses

- the first 10 parameters from the PCA model
- the first 10 from  $\mathbf{P}_{sym}$
- the first 10 from  $\mathbf{P}_{asym}$
- the first 5 from  $\mathbf{P}_{sym}$  and 5 from  $\mathbf{P}_{asym}$

A linear SVM is trained with 10-fold cross-validation. As can be seen from Table 5.3, symmetry parameters have slightly better accuracy than that of PCA model, while asymmetry parameters alone are worse in age classification than that of PCA model. This implies that symmetry information is more discriminative in age classification than asymmetry information. Since asymmetry variation is relevant to the age (faces in age group 0-11 are more symmetrical than that in  $> 60$  age group), the asymmetry parameters are still valid in age classification. The combination of 5  $\mathbf{P}_{sym}$  and 5  $\mathbf{P}_{asym}$  are better in age

---

classification than PCA model and  $\mathbf{P}_{sym}$  alone, which implies that the asymmetry information can improve the performance in age classification when combined with symmetry parameters.

## 5.6 Summary

We proposed a Symmetry-plane aware Coherent Point Drift (SA-CPD) algorithm and evaluated it on 3D images of the human head. This deformation method mitigates the tangential sliding problem seen in competing morphing algorithms, sometimes significantly, thereby improving the correspondence quality. The proposed method is also robust to outliers, missing data and Gaussian noise. The constructed morphable model based on the proposed deformation method has the best performance in both gender and age SVM-based classification compared to the leading competing methods. The deformation method is applicable to any shape sets that exhibit bilateral symmetry over a reflective symmetry plane.

We presented a craniofacial symmetry-factored statistical modelling pipeline and applied it to a craniofacial dataset of 1212 subjects. Contributions included Laplace-Beltrami regularized symmetrisation, symmetry-plane aware GPA and symmetry-factored statistical modelling method. We used it to build the first morphable model that makes craniofacial asymmetry explicit. A comprehensive evaluation shows that the proposed model has significantly better performance than standard PCA models, especially in terms of generalisation error and in the completion of head data. Symmetry-plane aware GPA can improve the performance of the standard PCA model. Additionally, we found that symmetry information is more discriminative than asymmetry information in age classification. Future work should focus on region-based asymmetry modelling and give a comprehensive study of facial symmetry/asymmetry in 3D face recognition.



# Chapter 6

## Modelling and Merging High Resolution Ear Shape

The shape of the ear has long been recognised as a means of biometric identification. There are many existing ear recognition systems and ear biometrics continues to be an active research area [60]. Morphable models provide powerful statistical priors on shape and so can be used in biometric ear analysis. However, it is very difficult to capture the detailed structure of the ear when morphing over the whole head. The high frequency detail of the fleshy folds is not captured and we are unable to construct a powerful statistical prior in the ear region. Thus we choose to model this in a separate process, which then generates the problem of having to merge separate parts together, namely the ear and the remainder of the head.

We present a pipeline capable of building a 3D Morphable Model (3DMM) of the human ear from a very limited training sample of 3D ears, using data augmentation. Deep learning can be used on 3D meshes, but a large-scale dataset is needed for training and it takes a huge effort to capture a sufficiently large number of 3D meshes. Since 3D data augmentation can generate new 3D data, it satisfies the need for a large dataset and has become an active research field.

We have 20 high quality 3D meshes of the ear [160], taken from 10 subjects, with the left ear reflected to be compatible with the right ear shape. This is insufficient to construct a 3D morphable model that is a good representation of the mean ear shape and the variance and covariances of size-and-shape (form), over a large population. However,

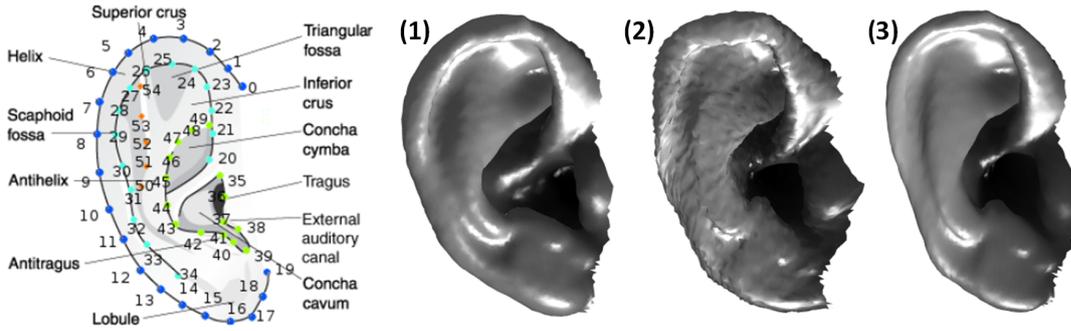


Figure 6.1: 55 landmarks on ear and their semantic annotations taken from [155]. Deformation: (1) template, (2) over-fitting, (3) deformed template.

with such a limited dataset, we construct an initial approximate model of the form:

$$\mathbf{X}(\alpha) = \bar{\mathbf{X}} + \mathbf{P}\alpha \quad (6.1)$$

where the 3DMM parameters are the mean shape  $\bar{\mathbf{X}}$ , the shape variation components  $\mathbf{P}$ , and shape parameters  $\alpha$ . The model has over 7K vertices (7111) and we employ a modified version of our morphing technique (Dai et al. [47]) to build the model. Subsequently, 3D data augmentation is able to generate new samples for the 3DMM construction, thereby boosting the initial morphable model in terms of its accuracy in representing larger populations. Recently, Zhou et al. [155] made a 2D ear image dataset available with 55 ground-truth landmarks and over 600 images, partitioned into 500 training images and 100 test images. Figure 6.1 (left) shows the 55 landmarks and their semantic annotations.

The chapter is structured as shown in Figure 6.2: in Section 6.1 we describe the ear model construction process including landmark-based 3DMM fitting, use of 3D deformation to overcome over-fitting, and 3D mesh manipulation regularized by 2D image information; in Section 6.2 we describe the process of ear merging with a 3DMM of full head. Section 6.3 describes the evaluation of both processes, while a final section is used for conclusions.

## 6.1 A 3DMM of the Ear

The process of data-augmented 3DMM construction is shown in Figure 6.3. Here, data augmentation has three stages: A) 3DMM fitting with 2D ear landmarks; B) 3D de-

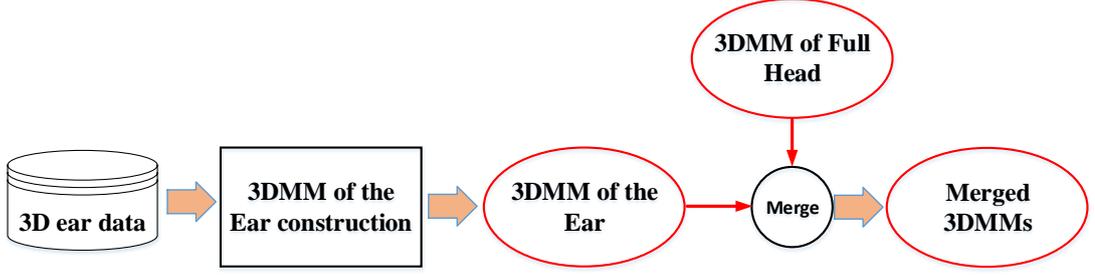


Figure 6.2: The flowchart of modelling and merging high resolution ear shape.

formation to overcome the over-fitting of the initial approximate 3DMM; C) 3D mesh manipulation to manipulate the projection of the landmarks in the augmented 3D mesh towards the manually-labelled 2D landmark positions. These three stages are described in the following subsections, with the iterative loop for model construction described. Our contributions are (i) the data-augmented model building pipeline and (ii) the first publicly-available 3DMM of the ear. This model is shown in Figure 6.4. In addition to the model, the augmented training data is made publicly available.

### 6.1.1 Landmark-based 3DMM Fitting

The scaled orthographic projection (SOP) [14] model assumes that variation in depth over the object is small relative to the mean distance from camera to object. Under this assumption, the projected 2D position of a 3D point  $\mathbf{X}_i = [x_i, y_i, z_i]^T \in \mathbb{R}^3$ , given by  $SOP(\mathbf{X}_i; \mathbf{R}, \mathbf{t}, s) \in \mathbb{R}^2$  does not depend on the distance of the point from the camera, but only on a uniform scale  $s$  given by the ratio of the focal length of the camera and the mean distance from camera to object:

$$SOP(\mathbf{X}_i; \mathbf{R}, \mathbf{T}, s) = s\mathbf{P}_o (\mathbf{R}\mathbf{X}_i + \mathbf{T}) \quad (6.2)$$

where the 3D pose parameters are given by a rotation matrix  $\mathbf{R}$  *s.t.*  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ ,  $\mathbf{R}^T \mathbf{R} = \mathbf{I}_3$  and 3D translation  $\mathbf{T} \in \mathbb{R}^3$ .  $\mathbf{P}_o$  is the orthogonal projection from 3D to 3D defined by

$$\mathbf{P}_o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

and so, defining the 2D translation,  $\mathbf{t}$  in the image plane we have

$$SOP(\mathbf{X}_i; \mathbf{R}, \mathbf{t}, s) = s\mathbf{P}_o \mathbf{R}\mathbf{X}_i + \mathbf{t}, \quad \mathbf{t} = s\mathbf{P}_o \mathbf{T} \quad (6.3)$$

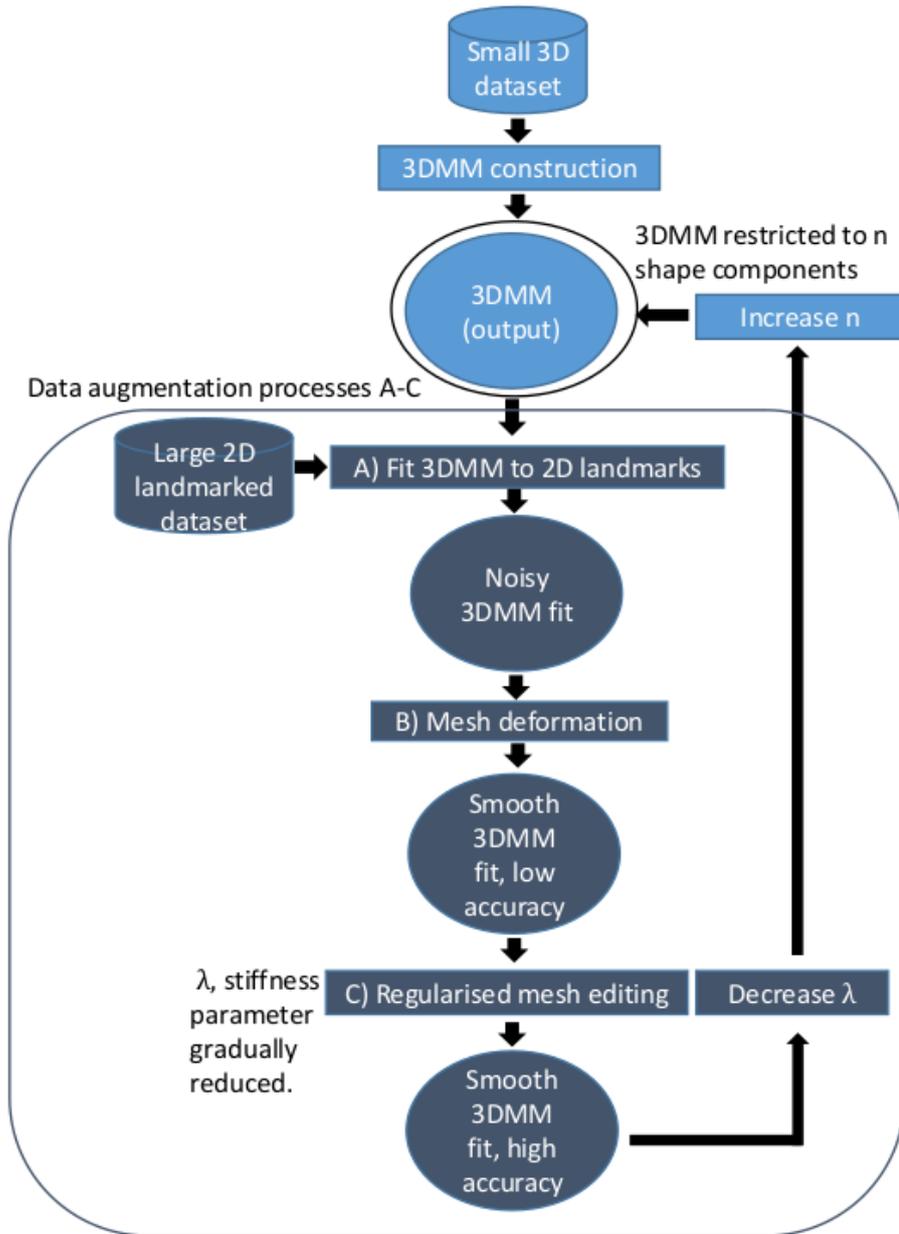


Figure 6.3: Iterative model construction process: A) 3DMM fitting with 2D ear landmarks; B) 3D deformation to overcome the over-fitting of the initial approximate 3DMM; C) 3D mesh manipulation to manipulate the projection of the landmarks in the augmented 3D mesh towards the manually-labelled 2D landmark positions.

We begin by showing how to fit a morphable model to  $M$  observed 2D positions  $\mathbf{x}_i = [u_i, v_i]^T (i = 1 \dots M)$  arising from the *SOP* projection of corresponding vertices in the morphable model. Without loss of generality, we assume that the  $i$ -th 2D position corresponds to the  $i$ -th vertex in the morphable model. The objective of fitting a morphable

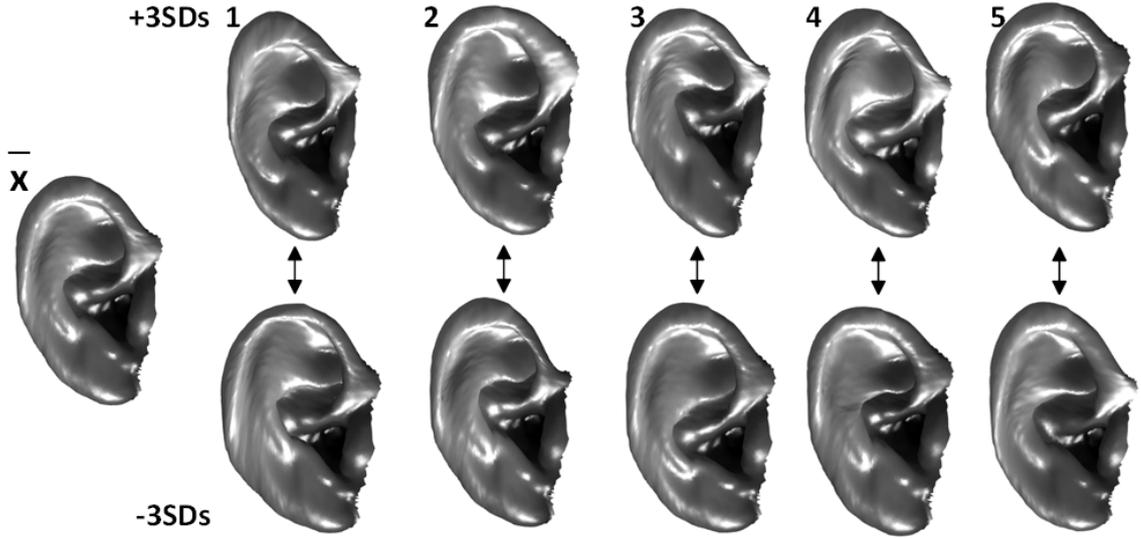


Figure 6.4: 3D morphable model of ear. The mean and the first five principal components are shown at +3SDs (top row) and -3SDs (bottom row).

model to these observations is to obtain the shape and pose parameters that minimise the reprojection error,  $E_{lmk}$ , between observed and predicted 2D landmark positions:

$$E_{lmk}(\alpha, \mathbf{R}, \mathbf{t}, s) = \frac{1}{M} \sum_{i=1}^M \|\mathbf{x}_i - SOP(\bar{\mathbf{X}}_i + \mathbf{P}_i \alpha; \mathbf{R}, \mathbf{t}, s)\|^2 \quad (6.4)$$

The problem is non-linear least squares that can be solved by various means. Here we use the trust region approach [39] encapsulated in Matlab's `lsqnonlin` function.

### 6.1.2 3D Mesh Deformation

The number of training subjects for the initial 3DMM is insufficient, so the 3DMM fitting to a 2D image with landmarks, described in Sec. 6.1.1 is over-fitted, appearing as surface noise, see Figure 6.1 (2). To overcome this, we employ the mean of the initial 3DMM, see Figure 6.1 (1) as a template, and we deform it using the Coherent Point Drift (CPD) algorithm [100] applied with a non-rigid deformation model, followed by a projection to corresponding points that is regularised by the template shape-preserving Laplace-Beltrami (LB) operator. The motivation for the deformation process is that the deformed template is able to preserve the same shape, the same number of vertices and also the same triangulation relationship as the over-fitted data, while it can overcome the noise due to over-fitting. The deformation algorithm works well because there is a known

one-to-one correspondence between the 7111 vertices on the template and the 7111 vertices on the target. The outcome is shown in Figure 6.1 (3).

### 6.1.3 Mesh Manipulation Regularised by 2D Landmarks

The locations of the fitted landmarks after the initial 3DMM fitting and the template deformation are not accurate, relative to the manually-labelled 2D landmarks. We overcome this by treating the template mesh manipulation as a mesh manipulation problem with two ingredients. First, position constraints are provided by those 2D landmarks, the correspondences of which are known in 3D mesh. Second, regularisation constraints are (again) provided by the LB operator, which acts to retain the local structure of the mesh.

The LB mesh manipulation problem can be written as a linear system of equations. Given the vertices of a 3D mesh stored in the matrix  $\mathbf{X} = [x_1, y_1, z_1, \dots, x_N, y_N, z_N]^T \in \mathbb{R}^{3N}$  and the 2D landmarks stored in the matrix  $\mathbf{x} = [u_1, v_1, \dots, u_M, v_M]^T \in \mathbb{R}^{2M}$ , we define the selection matrices  $\mathbf{S} \in [0, 1]^{3M \times 3N}$  that select the  $M$  vertices which are the correspondences of the 2D landmarks. This linear system can be written as:

$$\begin{pmatrix} \lambda \mathbf{L}_3 \\ \mathbf{G}(M)\mathbf{S} \end{pmatrix} \mathbf{X}_{\text{edit}} = \begin{pmatrix} \lambda \mathbf{L}_3 \mathbf{X} \\ \mathbf{x} \end{pmatrix} \quad (6.5)$$

$$\mathbf{G}(M) = \mathbf{I}_M \otimes \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

where  $\mathbf{I}_M$  is the  $M \times M$  identity matrix and  $\mathbf{G}(M) \in \mathbb{R}^{2M \times 3M}$  project the 3D landmarks to 2D,  $\mathbf{L}_3 \in \mathbb{R}^{3N \times 3N}$  is the cotangent Laplacian approximation to the LB operator [121] and  $\mathbf{X}_{\text{edit}} \in \mathbb{R}^{3N}$  are the edited vertex positions that we wish to solve for. The parameter  $\lambda$  weights the relative influence of the position and shape regularisation constraints, effectively determining the template shape ‘stiffness’ of the mesh manipulation process. As  $\lambda \rightarrow 0$  (reducing shape stiffness) the projected 3D landmarks in  $\mathbf{X}_{\text{edit}}$  tend towards the same positions as the 2D manual landmarks.

### 6.1.4 Similarity Alignment & Statistical Modelling

The collection of the augmented meshes are subjected to Generalised Procrustes Analysis (GPA) [72] to remove similarity effects (rotation, translation, scale), leaving only shape

information. (Scale cannot be included as we have no notion of scale within the 2D image dataset.) The aligned meshes are then subject to Principal Component Analysis (PCA), generating a 3DMM as a linear basis of shapes. This allows for the generation of novel shape instances.

### 6.1.5 3DMM Bootstrapping

We propose a 3DMM bootstrapping procedure where, at each bootstrap iteration, we rebuild the 3DMM and reapply it to the augmented dataset for an improved fitting to that dataset, and hence we can generate a better 3DMM in the next iteration. This approximate-to-accurate iterative system encapsulates each of the three key stages in Sec. 6.1.1 to Sec. 6.1.3 within each iteration. We push each procedure harder relative to the previous iteration, as follows: 1) we increase the number of the shape components in Sec. 6.1.1 to give the algorithm more variance to do the fitting; 2) we decrease  $\lambda$  in Sec. 6.1.3 to manipulate the projection of the landmarks in  $\mathbf{X}_{\text{edit}}$  towards the 2D landmarks position. 3DMM fitting and mesh manipulation are potentially fragile processes when the 3DMM is approximate, thus we push the algorithm step-by-step in this iterative fashion.

## 6.2 Merging 3D Morphable Models

Given that it is desirable to create a separate ear model to capture its detailed shape, we then have the problem of replacing the ear shape on the full head model with the improved ear model. As shown in Figure 6.5, the proposed framework for this operation includes two stages: 1) rigid alignment: rigidly align an ear sample randomly generated from the ear model to the ear part of the head sample that is randomly generated by head model; 2) mesh manipulation with Poisson field: typically there is a gap between the two samples, so it needs an extra transformation to merge the two meshes. These two processes are detailed in the following subsections.

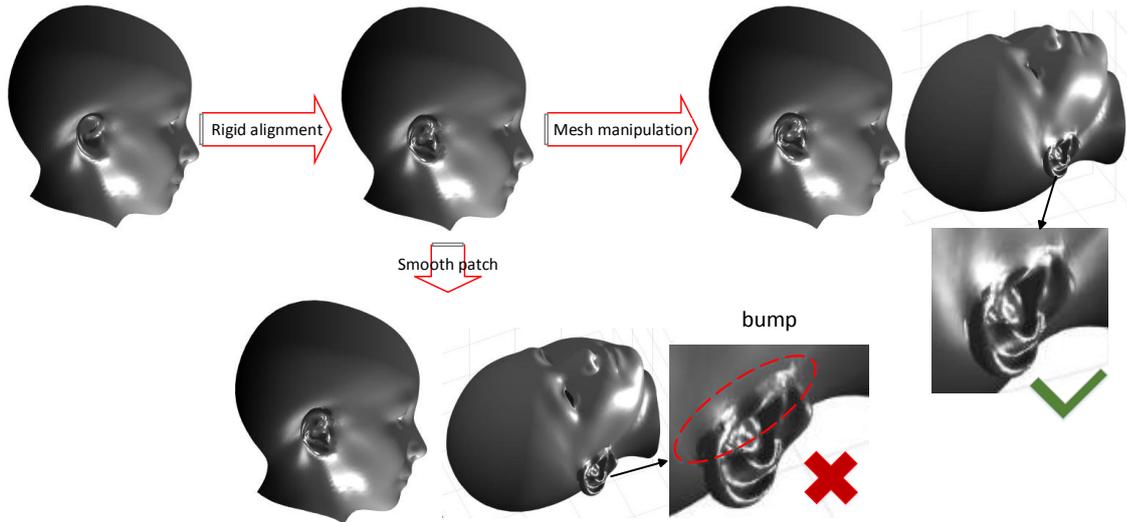


Figure 6.5: The merging flowchart includes two stages: 1) rigid alignment; 2) mesh manipulation. If the output of the rigid alignment undergoes a patch smoothing operation only, it suffers from a discontinuity problem, ending up with a bump. The mesh manipulation overcomes this.

### 6.2.1 Joint Template

The high resolution ear template replaces the ears on the head template, ending up with a joint template: a head template with a high resolution ear. We use the software package *meshmixer* [116] to blend a high resolution ear template with the head template to create our joint template. In other words, the triangulation relationship is already known. The problem that we need to solve is how to merge these two morphable models. Figure 6.6 demonstrates the difference between the triangulation of the original template and that of the joint template. We cut off the high resolution ear from the joint template. The connection relation between the two separate parts are known from the joint template. Then we can use this connection relation to merge the two separate morphable models. The key in this process is to solve the correct position and matched edge for the high resolution ear shape. The red points represent the shared vertices for the high resolution ear  $\mathbf{X}'_e$  and the rest of the head  $\mathbf{X}^*_{h-e}$ . When splitting the joint template into two parts:  $\mathbf{X}'_e$  and  $\mathbf{X}^*_{h-e}$ , the correct position of the red points on the ear part  $\mathbf{X}'_e$  should be the same position of the red points on the no-ear head part  $\mathbf{X}^*_{h-e}$ . Now these red points are called a ‘Poisson field’. Given the selection matrix  $\mathbf{S}'_{pf}$  as those that select the Poisson field on

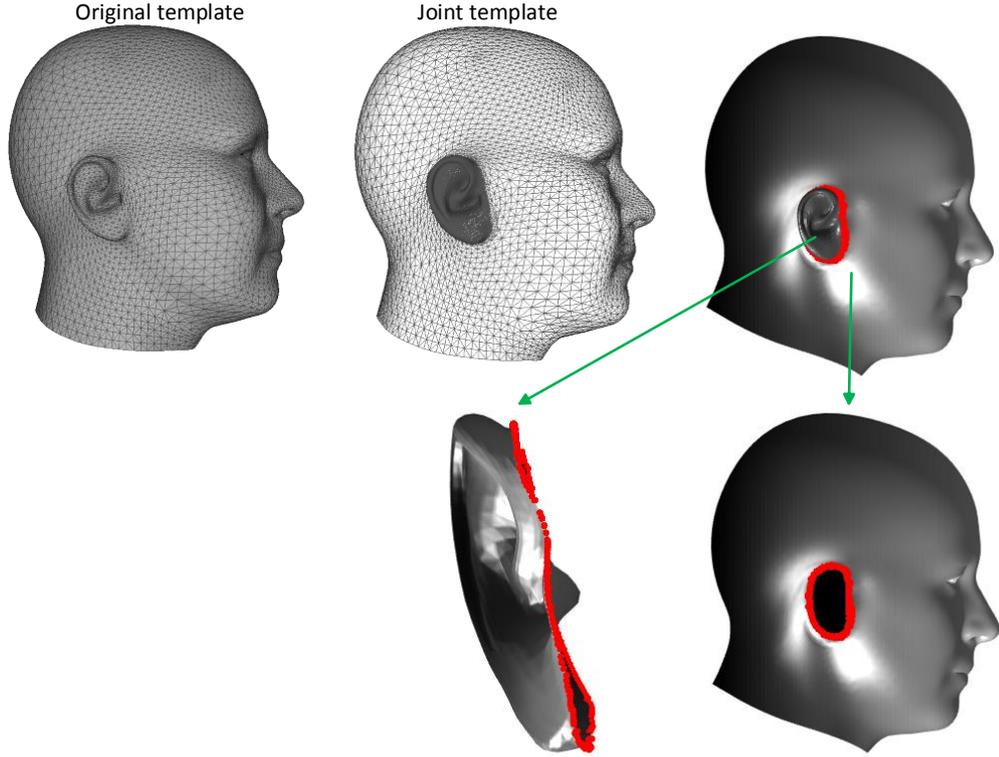


Figure 6.6: Joint template.

the high resolution 3D ear part  $\mathbf{X}'_e$ , this Poisson field can be represented as  $\mathbf{S}'_{\text{pf}}\mathbf{X}'_e$ . Given the selection matrix  $\mathbf{S}^*_{\text{pf}}$  as those that select the red points on the no-ear head part  $\mathbf{X}^*_{h-e}$ , these red points can be represented as  $\mathbf{S}^*_{\text{pf}}\mathbf{X}^*_{h-e}$ . For a sample generated by the part-based morphable model which merge the two models, it should satisfy the requirement:

$$\mathbf{S}'_{\text{pf}}\mathbf{X}'_e = \mathbf{S}^*_{\text{pf}}\mathbf{X}^*_{h-e} \quad (6.6)$$

### 6.2.2 Rigid alignment

We begin by showing how to rigidly align the high resolution 3D ear sample  $\mathbf{X}'_e$  to the low resolution 3D ear  $\mathbf{X}_e^*$  on the head mesh sample  $\mathbf{X}^*$ . Given a selection matrix  $\mathbf{S}_e^*$  as those that select the low resolution 3D ear, we can form  $\mathbf{X}_e^* = \mathbf{S}_e^*\mathbf{X}^*$ . We aim to find the rigid transformation matrix that can rigidly transform  $\mathbf{X}'_e$  to  $\mathbf{X}_e^*$ . This can be solved by normalising the scale of the high-resolution ear to that of the low-resolution ear, and then using ICP [18].

### 6.2.3 Mesh Manipulation with Poisson Field

These rigid transformation matrices are used for aligning the decomposed parts to the 3D raw data. The rigidly transformed facial parts tell the original template where it should be. We treat this as a mesh manipulation problem. We use Laplace-Beltrami mesh manipulation to manipulate the Poisson field on  $\mathbf{X}'_e$  towards the Poisson field on  $\mathbf{X}^*_e$  and the rest of  $\mathbf{X}'_e$  is moved ‘*as rigid as possible*’ along with this manipulation, as follows: (1) the *Poisson field* (pf) of the high resolution 3D ear mesh  $\mathbf{X}'_e$  is manipulated towards their target positions - red points in Figure 6.6; (2) all other parts of  $\mathbf{X}'_e$  are moved as rigidly as possible.

Given a refined high resolution 3D ear mesh, whose vertices are stored in the matrix  $\mathbf{X}'_{\text{refine}} \in \mathbb{R}^{p \times 3}$ . This linear system can be written as:

$$\begin{pmatrix} \mathbf{L} \\ \lambda \mathbf{S}'_{pf} \end{pmatrix} \mathbf{X}'_{\text{refine}} = \begin{pmatrix} \mathbf{L} \mathbf{X}'_e \\ \lambda \mathbf{S}^*_{pf} \mathbf{X}^*_{\text{h-e}} \end{pmatrix} \quad (6.7)$$

where  $\mathbf{L} \in \mathbb{R}^{p \times p}$  is the cotangent Laplacian approximation to the LB operator [121] and  $\mathbf{X}'_{\text{refine}}$  is the refined ear position that we wish to solve for. The parameter  $\lambda$  weights the relative influence of the position and regularisation constraints, effectively determining the ‘stiffness’ of the mesh manipulation. As  $\lambda \rightarrow 0$ , the ear part stays in its original position. As  $\lambda \rightarrow \infty$ , the Poisson field on the ear part is moved exactly onto its target positions - i.e. the red points on  $\mathbf{X}^*$ . After this mesh manipulation, the small artefacts can be removed by a simple patch smoothing technique proposed by Desbrun et al. [55], which employed an implicit integration method along with a scale-dependent Laplacian operator and a robust curvature flow operator to portray a smooth surface.

Given a selection matrix  $\mathbf{S}^*_{\text{h-e}}$  that selects the no-ear head part  $\mathbf{X}^*_{\text{h-e}}$  on the a head sample  $\mathbf{X}^*$ , which is generated by the head model, a new instance  $\mathbf{X}'$  generated by the merged model can also be represented as  $\mathbf{X}' = [\mathbf{X}'_{\text{refine}}; \mathbf{S}^*_{\text{h-e}} \mathbf{X}^*]$ .  $\mathbf{X}'_{\text{refine}}$  can be solved from a linear system and  $\mathbf{S}^*_{\text{h-e}} \mathbf{X}^*$  can be obtained from the head model linearly. So this part-based morphable model is still a linear PCA model, which facilitates its application in 3DMMs fitting to 2D images.

## 6.3 Evaluation

We used the proposed method to build a 3DMM of the ear over 500 training images and used the remaining 100 images for testing the performance in the given dataset [155]. There is no public 3DMM of the ear available for direct comparison. However, in section 6.3.1 and section 6.3.2 we evaluate the performance of model construction, both qualitatively and quantitatively, for several variants of our method. These include: i) the proposed method, using several bootstrapping iterations, and 500 data augmentation images, ii) the proposed method without any bootstrapping iterations (i.e. one pass of the three steps in Sec. 6.1) and 500 data augmentation images, iii) the initial 20-image 3DMM passed through the three steps in Sec. 6.1, with no data augmentation (Initial-v1 method) and iv) the initial 20-image 3DMM with just 3DMM fitting, i.e. no template morphing or mesh manipulation stages, and no data augmentation (Initial-v2 method). The performance improvement of ear model is presented in section 6.3.3. We compare the proposed ear merging method with other methods in section 6.3.4. The merged 3DMMs are presented in section 6.3.5.

### 6.3.1 Qualitative Evaluation

In order to validate the effectiveness of each step in augmentation, we demonstrate the outcome of each step in Figure 6.7. The landmark positions of the fitted results get closer and closer to the landmark ground-truth. Due to insufficient data, there exists obvious over-fitting from the 3DMM fitting to 2D landmarks at the beginning. The CPD-nonrigid deformation removes the noise, but it still keeps the same landmark position as the first step. The outcome of final step (LB mesh manipulation) is almost the same landmark positions as the ground-truth. This implies that the LB mesh manipulation refines the landmark position. Quantitative evaluation is included in next subsection. In order to demonstrate the accuracy of the ear shape, we rigidly align the 3D ear shape to the 2D landmarks on the 2D image. Then the ear shape with texture is cropped. As can be seen in Figure 6.8, there are almost no other pixels other than ear texture in the images of the 3D ear with texture, which implies that the end result is very accurate. The proposed method can handle different head poses. After mesh manipulation, the projected positions of the

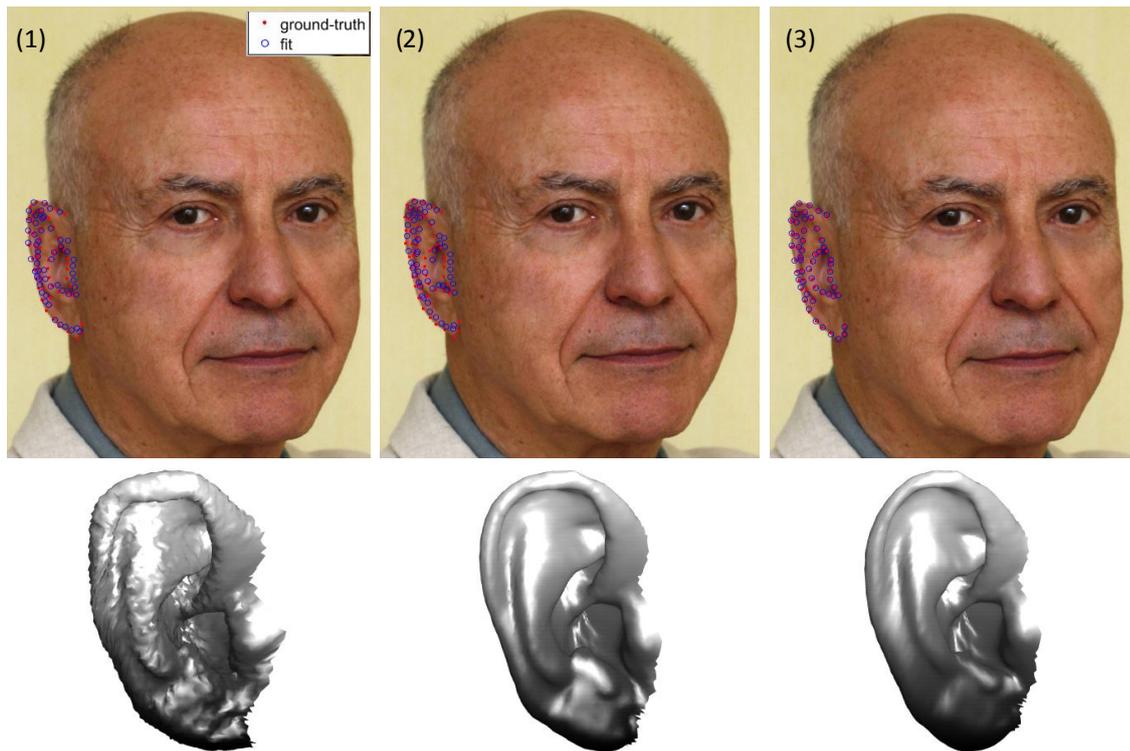


Figure 6.7: Results of each step in data augmentation: (1) 3DMM fitting to 2D landmarks; (2) deformation to overcome over-fitting; (3) LB mesh manipulation to refine the landmark position.

model's 3D landmarks are almost the same as the ground-truth 2D landmarks.

### 6.3.2 Quantitative Evaluation

We use two metrics: landmark error and fitting consistency to evaluate the performance quantitatively.

**Landmark Error:** This measure is calculated by the average landmark distance error between the projected 3D landmarks and the 2D landmarks, over the test set (100 images). As shown in Figure 6.9(1), the proposed method has the lowest landmark error. Initial-v1 indicates the results with data augmentation. When compared with no data augmentation version (only 3DMM fitting to 2D images), the mean of landmark error decreases from 3.7 to 2.2, which implies that the data augmentation improves the accuracy of ear shape generation. The bootstrapping stage makes the accuracy of ear shape generation even better.



Figure 6.8: Augmentation results: 1st row - original images, 2nd row - 3D landmarks projection to 2D images, 3rd row - augmented data with texture.

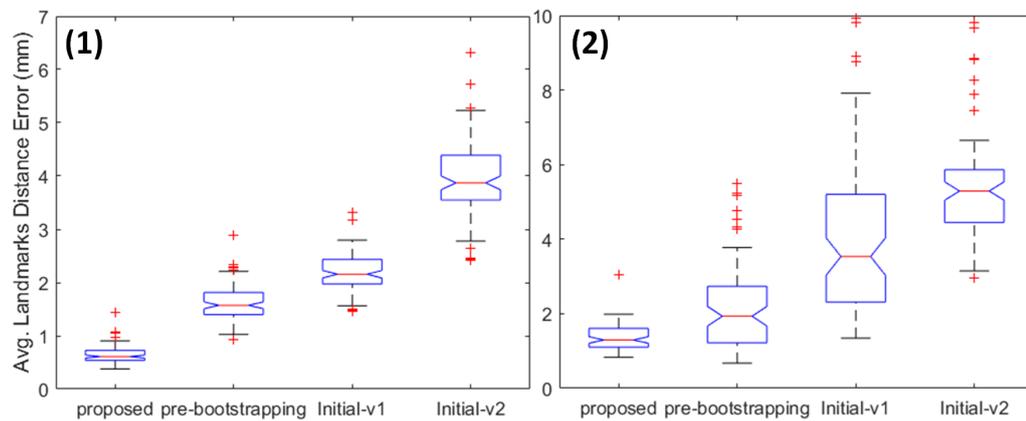


Figure 6.9: Average landmarks distance error for four system variants: (1) Landmark error, (2) Fitting consistency.

**Fitting Consistency:** The dataset contains multiple images of the same person, as shown in the first two columns of Figure 6.8. This allows us to do some consistency

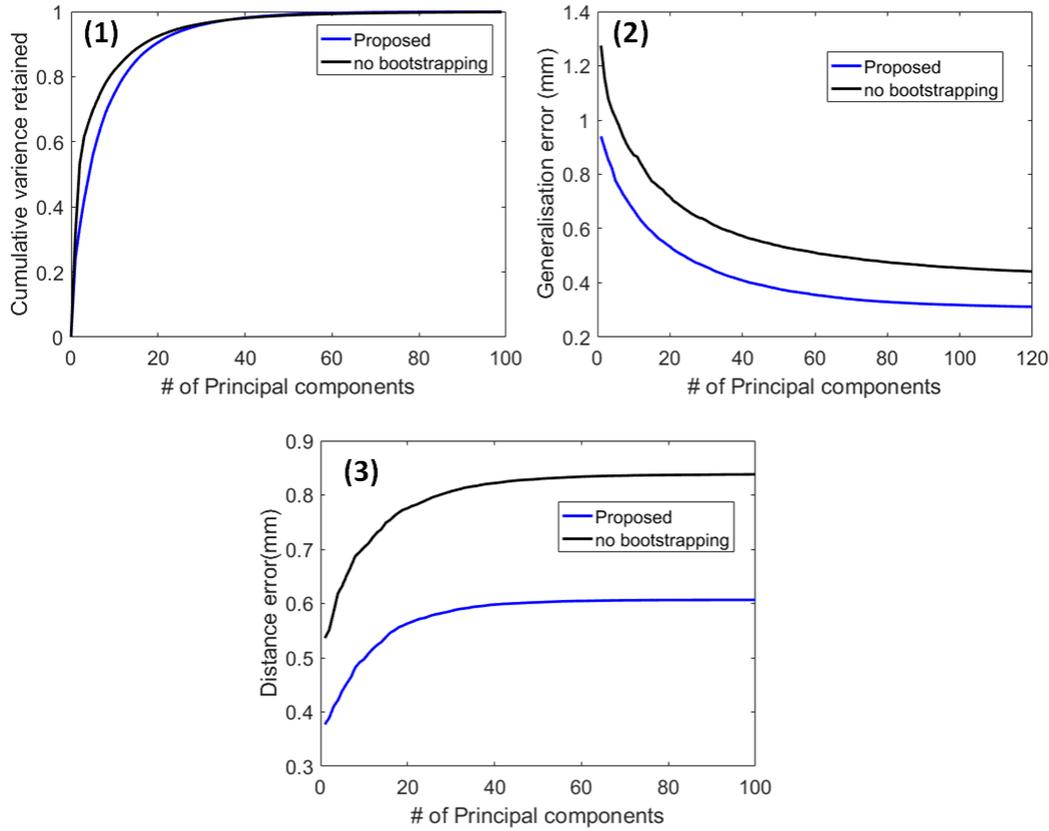


Figure 6.10: Model evaluation: (1) Compactness, (2) Generalisation, (3) Specificity.

checking in the following way. First we fit the 3D model to the first image of a pair, thus fixing the 3D model shape. Then, without changing the model shape, we project it into the second image and measure the mean landmark error relative to the manual 2D landmarks. We compensate for differences in scale between the two images in the fitting process. As shown in Figure 6.9(2), the proposed method has the lowest distance error, which implies that the fitting from the proposed method is more consistent with the other images of the same person.

### 6.3.3 Ear Model Evaluation

Since model evaluation requires that the models should have the same number of principal components, we compare the proposed method and the proposed method without bootstrapping. For quantitative model evaluation, Styner et al. [123] give detailed descriptions of three metrics: compactness, generalisation and specificity. The compactness of the

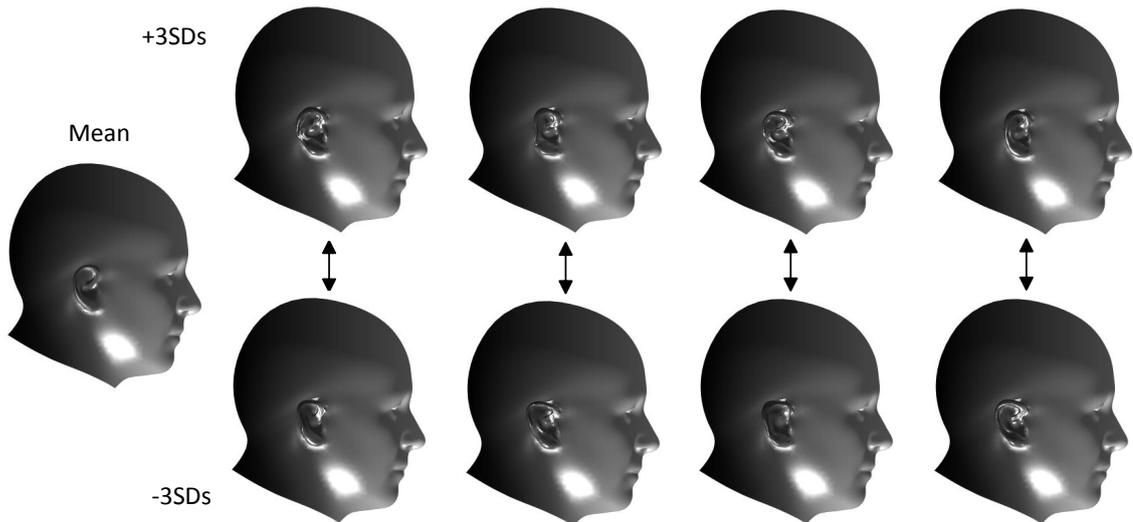


Figure 6.11: The first 4 modes of high resolution 3DMM of ear merged into a mean head from LYHM.

model describes the number of parameters required to express some fraction of the variance in the training set, fewer is better. As can be from Figure 6.10, the proposed method without bootstrapping has better compactness than the proposed method when  $< 25$  principal components are used. When  $> 25$  principal components are used, the compactness is similar. The proposed method has the lower generalisation error, which implies that proposed method has the better performance in describing unseen examples. Specificity measures how well a model is able to generate instances that are similar to real data. The proposed method has the lower distance error, which implies that the proposed method is better at generating instances close to real data.

#### 6.3.4 Comparison of Ear Merging

We compare the proposed ear merging method with mesh smoothing [55] after ear alignment and Laplacian mesh manipulation [122]. As shown in Figure 6.13 (2), with ear alignment, the high resolution ear mesh is rigidly transformed to the right position. If we use mesh smoothing directly after ear alignment, the joint area ends up with bumps which is presented in Figure 6.13 (3). It is desirable to have smoothed joint area for ear merging. Laplacian mesh manipulation is based on the Laplacian to do interactive free-form deformation. As can be seen from Figure 6.13 (4), Laplacian mesh manipulation presents

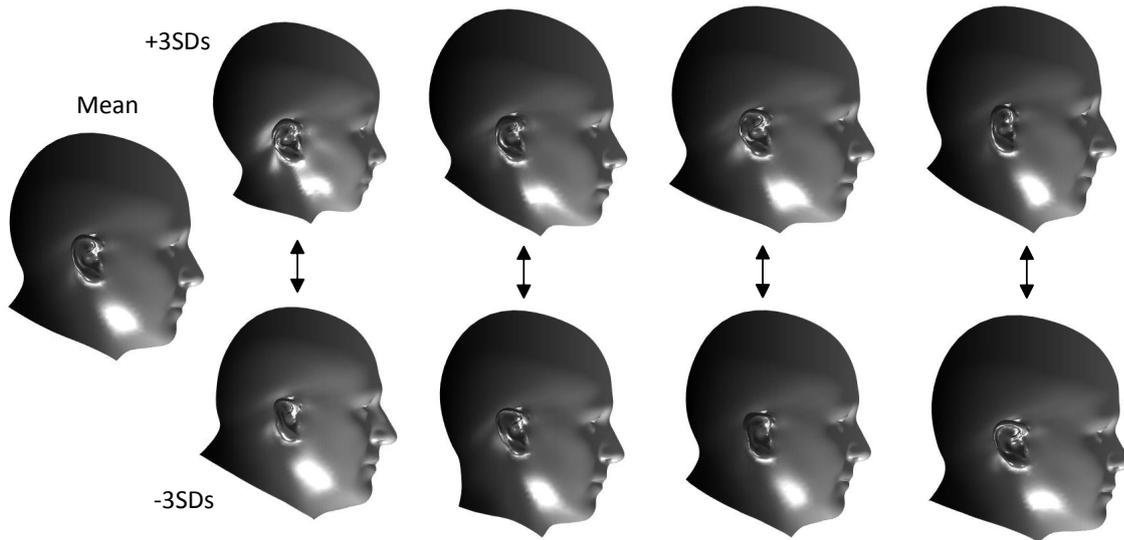


Figure 6.12: The first 4 modes of high resolution 3DMM of ear merged into the first 4 modes of head model.

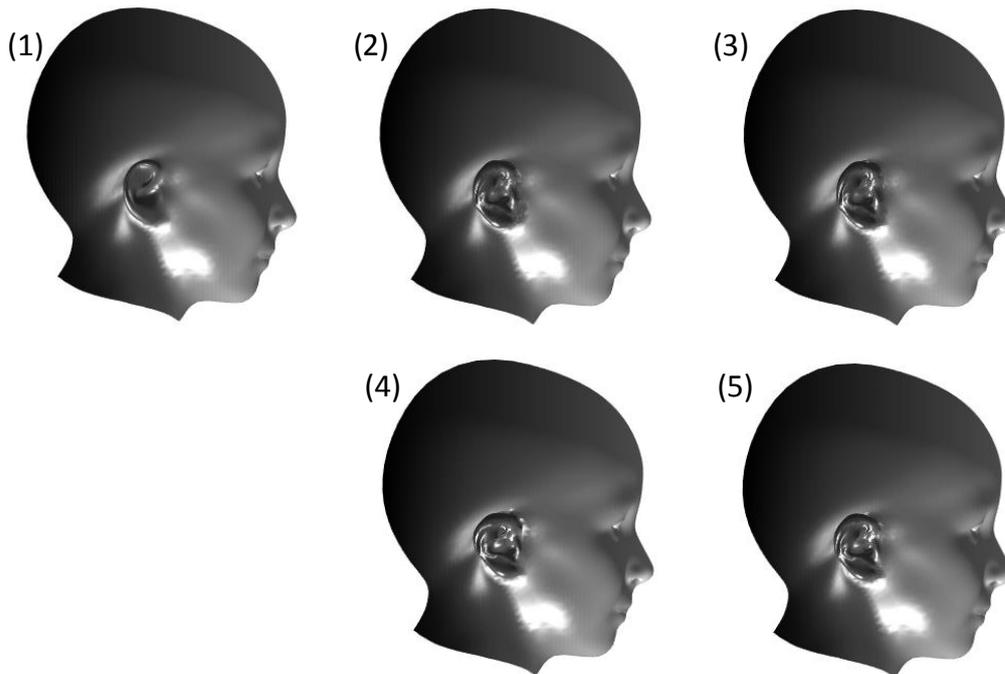


Figure 6.13: Comparison of ear merging: (1) original head mesh; (2) ear alignment; (3) mesh smoothing after ear alignment; (4) Laplacian mesh manipulation; (5) proposed.

a non-rigid deformation in ear shape. This changes the high resolution ear shape, which is not desirable in this process. Figure 6.13 (5) demonstrates the merging outcome of the proposed method. It shows a smoothed joint area between the ear part and face part. The

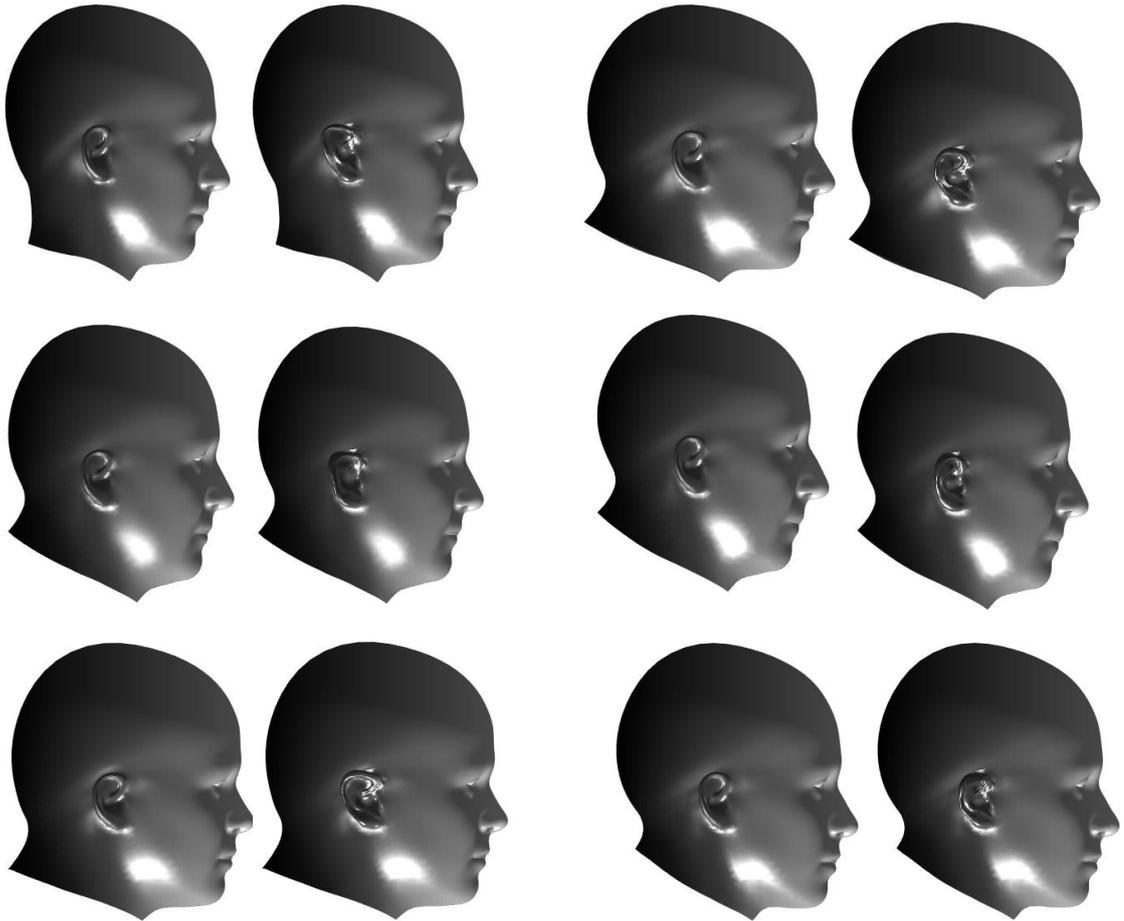


Figure 6.14: Ear merging results: every pair of images include one original head mesh and a merged result.

ear shape is the same as that after rigid alignment, which results from the ‘*as rigid as possible*’ property of LB mesh manipulation. Figure 6.14 shows the ear merging results for different identities using the proposed method.

### 6.3.5 Visualisation of the Merged Morphable Model

The merged morphable model is derived from merging the proposed ear model with a head model. Figure 6.11 demonstrates the first 4 modes of high resolution 3DMM of ear merged into a mean head from LYHM. In this case, the head shape is fixed and the ear shape is varied. Figure 6.12 presents the first 4 modes of high resolution 3DMM of ear merged into the first 4 modes of head model. Here, the head shape and ear shape are both varied. In order to validate the improvement in 3DMM fitting to 2D images,



Figure 6.15: Fitting results of a single 3DMM and the proposed part-based 3DMM.

we use the landmark fitting algorithm in iBug ear dataset [155]. The ear landmarks are given and we use a facial landmarking system from [157] for 3DMM fitting. The fitting results are shown in Figure 6.15. It shows that the part-based morphable model improves the performance of 3DMM fitting to 2D images when compared with a head model only. Actually, there is not much shape variation in the ear that is generated by the head model only.

## 6.4 Summary

We proposed an iterative 3DMM construction using 3D data augmentation to bootstrap a strong 3DMM of the human ear from a weak one. The method overcomes noisy overfitting and manipulates the projection of 3D landmarks towards the desired 2D landmark positions. Evaluation demonstrates that the method lowers the landmark error and the fitted data is more consistent within images of the same person. The bootstrapping strategy improves the model performance in both generalisation and specificity. The limitation is the requirement for manual 2D landmarks. We proposed a framework of merging high resolution ear shape with a 3DMM of the head. The merged morphable models provide more ear shape variation.



# Chapter 7

## Conclusions

In this chapter, we summarise the main contributions of the thesis in several areas of 3D morphable model construction. We then critically analyse weaknesses and finally present directions for further work.

### 7.1 Summary of Contributions

In the introduction and literature review, we highlighted the following facts: (i) most previous methods on morphable model construction are not fully automatic; (ii) craniofacial surgeons often like to employ a 2D profile view for operation outcome assessment; (iii) craniofacial symmetry properties are not considered in existing template morphing algorithms and statistical modelling techniques, and (iv) useful 3DMMs of the ear region demand more ear shape variation. Contributions to address each of these is presented in the following four subsections.

#### 7.1.1 Modelling of Orthogonal Craniofacial Profiles

We propose a new pipeline to build a 2D morphable model of the craniofacial sagittal profile and augment it with profile models from frontal and top down views. We also integrate all three profiles into a single model, thus capturing any correlations within and between the three profile shapes more explicitly and clearly than is possible with PCA analysis on a full 3D model.

### 7.1.2 3D Statistical Modelling Pipeline

We propose a fully automatic pipeline to build 3DMMs, with contributions in pose normalisation and dense correspondence. In particular we present a fully automatic registration framework using an adaptive template. We also present a high quality texture map from several views of cameras for the morphed template and use the texture map for texture modelling. We build both global craniofacial 3DMMs and demographic sub-population 3DMMs from more than 1200 distinct identities in the *Headspace* dataset [1,57] and we make them publicly available <sup>1</sup>. To our best knowledge, our models are the first public shape-and-texture craniofacial 3DMMs of the *full* human head. We demonstrate *flexibility modes* of our model such that, when given a fixed face shape, we compute the range of possible cranial shapes and vice-versa. We demonstrate the first clinical use of craniofacial 3DMMs in the assessment of two different types of surgical intervention applied to the crania. Finally, we describe a fully automatic way to do the craniofacial anthropometric measurements using the proposed 3DMMs. The results validate that the face and cranium grow significantly before 20 years old, and face grows more significantly than the cranium.

### 7.1.3 Symmetry-factored Statistical Modelling

We propose a Symmetry-aware Coherent Point Drift (SA-CPD) algorithm and evaluated it on 3D images of the human head. This deformation method mitigates the tangential sliding problem seen in competing morphing algorithms, sometimes significantly, thereby improving the correspondence quality. The proposed method is also robust against outliers, missing data and Gaussian noise. Based on the symmetry constrained correspondence output of SA-CPD, we present a symmetry-factored statistical modelling method for craniofacial shape. Our main contribution is to show how to build a statistical model with separate parameters for symmetric and asymmetric variations. The resulting model is still linear and so can be used in place of any existing 3DMM but with the additional ability to separate symmetric from asymmetric variation. This includes a method for symmetrisation regularised by the Laplace-Beltrami operator, symmetry-aware GPA and the

---

<sup>1</sup><https://www-users.cs.york.ac.uk/~nep/research/>

symmetry-factored statistical modelling method. Comprehensive evaluation shows that the proposed model has better performance than competing algorithms in the literature.

#### **7.1.4 Modelling and Merging High Resolution Ear Shape**

We present an iterative process of refinement for a 3D Morphable Model (3DMM) of the human ear that employs data augmentation. The process employs the following stages 1) landmark-based 3DMM fitting; 2) 3D template deformation to overcome noisy overfitting; 3) 3D mesh editing, to improve the fit to manual 2D landmarks. These processes are wrapped in an iterative procedure that is able to bootstrap a weak, approximate model into a significantly better model. Evaluations using several performance metrics verify the improvement of our model using the proposed algorithm. We use this new 3DMM model-booting algorithm to generate a refined 3D morphable model of the human ear, and we make this new model and our augmented training dataset public. We merge the proposed 3DMMs of ear with the full head model. This merged morphable model provides significantly more shape variation and shape detail of the ear than when the head is modelled as a single part.

## **7.2 Critical Analysis**

There are a number of angles from which criticisms may be levelled at the work presented in this thesis. Some of these weaknesses could be addressed by further work or by incorporating the techniques we have developed into other frameworks.

### **7.2.1 3D Facial Landmarking:**

The 3D facial landmarking in this thesis relies on texture coordinates to project 2D landmarks onto the 3D mesh. Even though there are landmarking systems detecting 3D landmarks from 2D images, the 3D landmarks detection is still limited by the ambiguity of 2D view. A fast and pose-invariant facial landmarks detection directly on 3D mesh is urgently needed.

### **7.2.2 Template Morphing:**

The proposed template morphing algorithms address the dense correspondence problem for craniofacial data with subjects having a neutral expression. It may end up with a poor morphing result when dealing with a strong facial expression. A more robust algorithm is required for data with strong and complex expressions, which would help a lot in dynamic 3D data registration. Another problem is that the current methods are not real-time applicable. It is very interesting problem to make real-time template morphing feasible for high resolution 3D data.

### **7.2.3 Symmetry-factored Modelling:**

The asymmetric variation is a small signal for PCA when compared with symmetric variation. Some high frequency signals which may be very interesting for researchers in different domains are easy to filter out using PCA. This is the same for all PCA-based statistical modelling methods.

### **7.2.4 Merging 3D Morphable Models:**

Merging multiple morphable models provides more shape variation for those local parts like the ear, nose and eye regions. However, the separate parts of the merged morphable models are still uncorrelated. In other words, they may generate an invalid face that does not exist in the real world. This incredible flexibility of shape variation is favored by the games and film industry, but it is less meaningful for shape analysis. The proposed merged morphable model bridges two separate models. The next generation should have the ability to turn the uncorrelated part-based model into a correlated one.

## **7.3 Future Work**

Throughout this thesis we have drawn attention to areas which warrant further research and routes by which the results presented may be improved upon. We conclude the thesis by providing clear directions for future research in 3DMMs.

### 7.3.1 3D Statistical Modelling Pipeline

Recent progress on geometric deep learning features a promising development in 3D mesh landmarking and correspondence establishment. One feasible way is to transform the 3D mesh into a UV map. Unlike a depth map, there is no spatial information loss in the UV map transformation. Then an auto-encoder can use these ingredients for landmarking and correspondence establishment. Once trained, the landmarking and correspondence establishment operating directly on a 3D mesh can be performed in real time. An auto-encoder also has the ability to replace PCA in the statistical modelling process. Auto-encoding with one layer is an analogue to PCA in statistical modelling. It is expected that more layers in an auto-encoder would help in explaining the shape variation and retaining high frequency signals.

The traditional methods like PCA need GPA to filter out similarity effects before modelling the shape variation. Moreover, two separate models are needed for shape and texture modelling. In future work, we aim at modelling the shape and texture variation using deep learning to achieve three goals: 1) learn a latent representation of shape variation which is invariant to similarity effects, 2) compose a six-channel input to model shape and texture variation at the same time, 3) use the 3D point sets as the input instead of a geometric image representation, such as the UV representation.

### 7.3.2 Merging 3D Morphable Models

The problem in our current part-based morphable model is that the separate parts are unrelated. In fact, both unrelated and related parts are required according to different applications. In order to make it correlated, we can generate synthetic data from 2D images using the current merged morphable model. Then we can build a new morphable model from the synthetic data. The vertex resolution is not uniform across the mesh, where weighted PCA can be applied to overcome this problem. In this way, the separate parts become correlated. However, this will end up with a global morphable model including all the shape variation in the separate 3DMMs. This may introduce an invalid shape variation to the 3DMM. How to solve this is the key to the the next version of merged morphable models.



# Appendix A

## List of Publications

- H Dai, N Pears, WA Smith, C Duncan, Symmetric Shape Morphing for 3D Face and Head Modelling. IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018).
- H Dai, N Pears, WA Smith, A Data-augmented 3D Morphable Model of the Ear. IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018).
- H Dai, N Pears, WA Smith, Non-rigid 3D Shape Registration using an Adaptive Template. Proceedings of the European Conference on Computer Vision Workshop (ECCV-W) Munich, Germany, 2018.
- H Dai, N Pears, WA Smith, C Duncan, A 3D Morphable Model of Craniofacial Shape and Texture Variation. Proceedings of the International Conference on Computer Vision (ICCV) 2017.
- H Dai, WA Smith, N Pears, C Duncan, Symmetry-factored Statistical Modelling of Craniofacial Shape. Proceedings of the International Conference on Computer Vision Workshop (ICCV-W) PeopleCap 2017.
- B Robertson, H Dai, N Pears, C Duncan, A morphable model of the human head validating the outcomes of an age-dependent scaphocephaly correction. International Journal of Oral and Maxillofacial Surgery.

- H Dai, N Pears, C Duncan, 2D Morphable Model of Craniofacial Profile and Its Application to Craniosynostosis. Annual Conference on Medical Image Understanding and Analysis (MIUA), 731-742 2017.
- H Dai, N Pears, C Duncan, Modelling of Orthogonal Craniofacial Profiles. Journal of Imaging 3 (4), 55, 2017.
- Y Ding, X Zhao, Z Zhang, H Dai, Image quality assessment based on multi-order local features description, modeling and quantification. IEICE Transactions on Information and Systems 100 (6), 1303-1315, 2017
- C Zhang, WAP Smith, A Dessein, N Pears, H Dai, Functional Faces: Groupwise Dense Correspondence using Functional Maps. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016.
- Y Zhao, Y Ding, G Yu, H Dai, Computer-Aided Detection of Microcalcifications Using Image Enhancement and Support Vector Classification. Journal of Medical Imaging and Health Informatics 6 (7), 1823-1827 2016.

# Appendix B

## Hierarchical Parts-based Template Morphing Framework

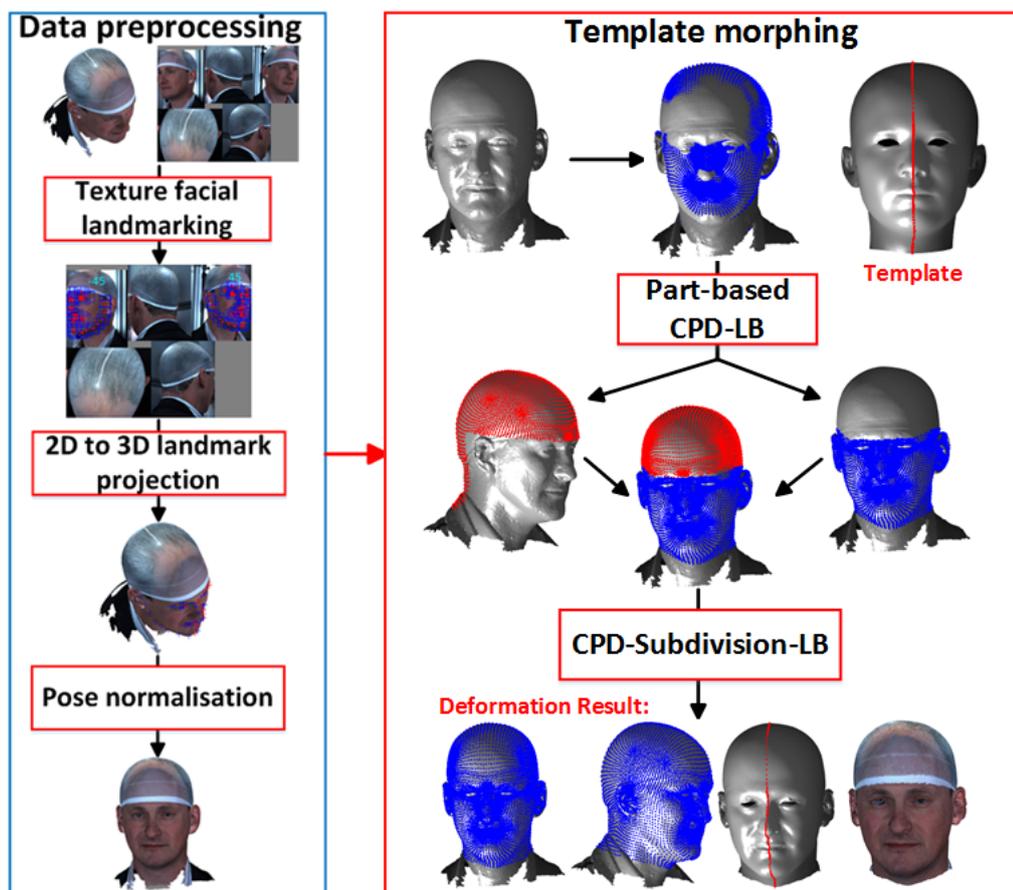


Figure B.1: Parts-based template morphing framework

### **B.0.1 Automatic 2D/3D Facial Landmarking**

We use the method of Zhu and Ramanan [157] to locate facial landmarks on the texture channel of each 3D image. This 2D image contains all 5 viewpoints of the capture system and usually two face detections are found, 15-45 degrees yaw from frontal pose, corresponding to the left and right side of the face. Detected 2D points are in a tree structure and are projected to 3D using OBJ texture coordinates.

### **B.0.2 Pose Normalisation**

Each face detection employs one of thirteen tree models [157] and we automatically learn how to orientate each of these to frontal pose, based on their 3D structure. To do this, we apply GPA to each collection of 3D trees (11 of the 13 models are used by the dataset) and find the nearest-to-mean tree shape in a scale-normalised setting. We then apply a 3D face landmarker [43] to the 3D data of the nearest-to-mean tree shape (11 of these), which generates a set of 14 landmarks with clear semantic meaning. Finally, we find the alignment that moves the symmetry plane of these 14 landmarks to the Y-Z plane with the nasion above the subnasale (larger Y coordinate) and at the same Z-coordinate, in order to normalise the tilt (X rotation). To complete the training phase, the mean 3D tree points for each of the 13 trees are then carried into this canonical frontal pose using the same rotation, and are used as reference points for the frontal pose normalisation of the 3D trees.

In around 1% of the dataset, only one tree is detected and that is used for pose normalisation, and in the rest 2-3 images are detected. In the cases where 3 trees are detected, the lowest scoring tree is always false positive and can be discarded. For the remaining two trees, a weighted combination of the two rotations is computed using quaternions, where the weighting is based on the mean Euclidean error to the mean tree, in the appropriate tree component.

## B.1 Dense Correspondence

In this section, we propose a new template morphing framework along with optical flow correspondence refinement. Both shape information and texture information are used for correspondence establishment.

### B.1.1 Template Morphing

Compared to NICP, CPD can avoid tangential point sliding and over-fitting, resulting in a good fit and good symmetry preservation. But standard CPD fails to resolve the under-fitting problem, where convergence terminates before good correspondence is established over the ears. This is because there are relatively few points on the ears and the facial/cranial points dominate. To avoid under-fitting, we propose a two-stage approach that consists of CPD followed by mutual nearest neighbour projection regularised by the Laplace-Beltrami (LB) operator. This two-stage approach is then applied within a parts-based framework.

**Two-stage CPD-LB Approach:** When CPD has deformed the template close to the scan, point projection is required to eliminate any (normal) shape distance error. Point projection is a potentially fragile process. If the scan data is incomplete or noisy then projecting vertices from the deformed template to their nearest vertex or surface position on the scan may cause large artefacts. We overcome this by treating the projection operation as a mesh editing problem with two ingredients. First, position constraints are provided by those vertices with mutual nearest neighbours between the deformed template and raw scan. Using mutual nearest neighbours reduces sensitivity to missing data. Second, regularisation constraints are provided by the LB operator which acts to retain the local structure of the mesh.

Thus we optimise two cost functions in a sequential manner. The Expectation-Maximization (EM) algorithm is used to solve the CPD cost function  $\mathbf{E}_{CPD}$ . For this cost function, we refer to [101]. For the second stage, we write the LB mesh editing problem as a linear system of equations. Given the vertices of a scan stored in the matrix  $\mathbf{X}_{scan} \in \mathbb{R}^{n \times 3}$  and the deformed template obtained by CPD whose vertices are stored in the matrix  $\mathbf{X}_{deformed} \in \mathbb{R}^{p \times 3}$ , we define the selection matrices  $\mathbf{S}_1 \in [0, 1]^{m \times p}$  and  $\mathbf{S}_2 \in [0, 1]^{m \times n}$

as those that select the  $m$  vertices with mutual nearest neighbours from deformed template and scan respectively. This linear system can be written as:

$$\begin{pmatrix} \lambda \mathbf{L} \\ \mathbf{S}_1 \end{pmatrix} \mathbf{X}_{\text{proj}} = \begin{pmatrix} \lambda \mathbf{L} \mathbf{X}_{\text{deformed}} \\ \mathbf{S}_2 \mathbf{X}_{\text{scan}} \end{pmatrix} \quad (\text{B.1.1})$$

where  $\mathbf{L} \in \mathbb{R}^{p \times p}$  is the cotangent Laplacian approximation to the LB operator and  $\mathbf{X}_{\text{proj}} \in \mathbb{R}^{p \times 3}$  are the projected vertex positions that we wish to solve for. The parameter  $\lambda$  weights the relative influence of the position and regularisation constraints, effectively determining the ‘stiffness’ of the projection. As  $\lambda \rightarrow 0$ , the projection tends towards nearest neighbour projection. As  $\lambda \rightarrow \infty$ , the deformed template will only be allowed to rigidly transform.

**Hierarchical Parts-based CPD-LB Framework:** Our template fitting algorithm, using temporary splitting of the template into parts and multiple applications of the two-stage CPD-LB deformation process is illustrated in the central panel of FigureB.1 and consists of the following six steps:

1. CPD-affine deforms the full-head template to fit the pose-normalised input until it converges to a global minimum according to the convergence criterion.
2. The template is then divided into two parts: cranial (red part in FigureB.1) and facial (blue part in FigureB.1) using a predetermined manual mesh segmentation. The input mesh is also divided into two corresponding parts using nearest neighbours.
3. The segmentation in step 2 releases the two individual parts from the global minimum in step 1 and CPD affine continues on the two parts separately to re-establish convergence. In practice, the cranial part is already very close to its global minimum, but the facial part continues to deform over many more iterations.
4. After convergence, the two parts of the template are updated using LB-regularised projection; however, there may exist a gap or an overlap between the two parts, due to the separate deformation processes.
5. The deformed full template from step 1 is now used to deform towards the two deformed parts-based templates from the previous step. Note that we now have a known one-to-one correspondence between the full template and the two template parts. Under these circumstances CPD-nonrigid performs well and is used.

6. After CPD-nonrigid converges, a final stage of LB-regularised projection is used to give the final morphed template of the full head, which is devoid of any overlaps or discontinuities that appear between the two separate parts in step 4 and is superior to the initial deformation from step 1 (standard CPD-affine).
7. Morphed template mesh subdivision followed by LB-regularised projection is used to upscale resolution.

## B.2 Results

We select 1212 individuals (606 males and 606 females) to evaluate our part-based template morphing framework.

### B.2.1 Correspondence Comparison

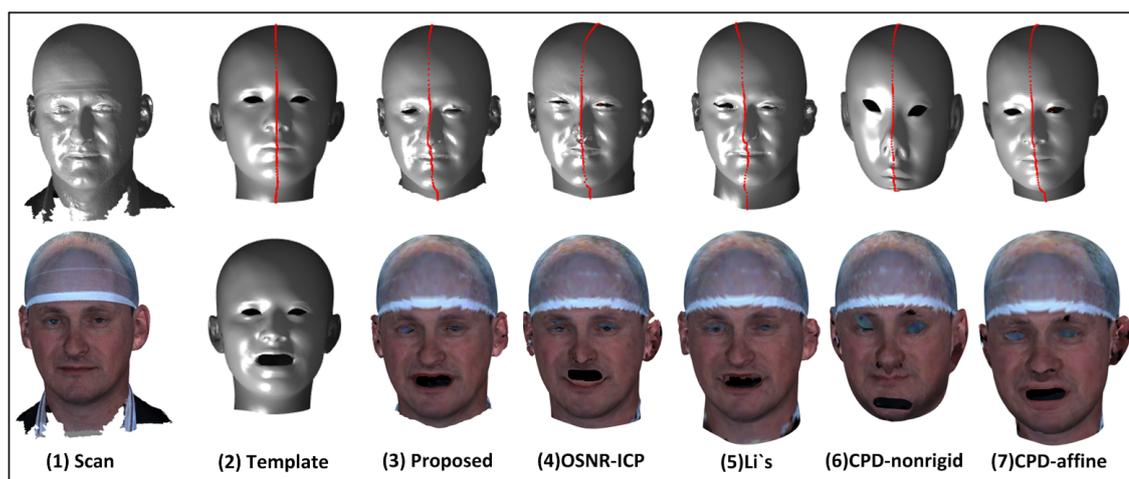


Figure B.2: Deformation results and their expression rendering with texture.

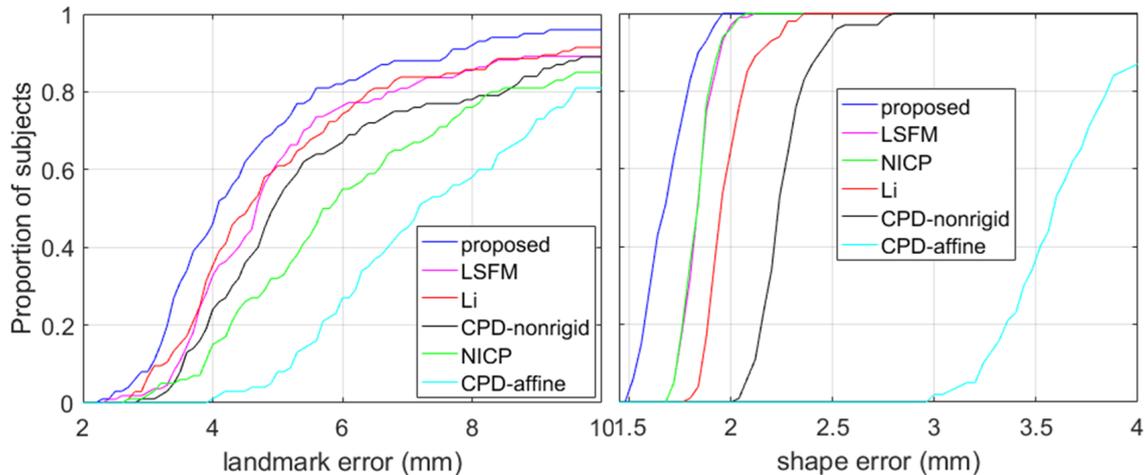


Figure B.3: Proportion of subjects with  $<$  Euclidean distance error: left-landmark error, right-shape error. The higher, the better.

**Qualitative Evaluation:** The proposed method is compared with NICP [8], LSFM pipeline [27], Li’s method [92] and standard CPD [101]. Expression rendering with texture is used to validate correspondence accuracy. Figure B.2 shows that the proposed method is qualitatively superior to other methods with respect to accuracy of the symmetry line, correspondence accuracy, and shape preservation relative to the input scan. Note that, the symmetry line slides to one side on the cranium for several methods due to over-fitting. NICP overfits in both the eye and mouth regions. Although, the (normal) shape error is low, the registered mouth is not in the correct position. Li’s method also overfits in the ear region. In contrast, CPD underfits that region in both affine and nonrigid versions. Figure B.2 shows that the proposed method with subdivision improves the quality of texture and captures more shape detail. But it also causes over-fitting in the eye region.

**Quantitative Evaluation:** We use 14 manual facial landmarks over 100 subjects to measure landmark error of each template deformation method. As shown in Figure B.3-left, 80% of landmark errors are less than 6 mm for our method. Figure B.3-right shows that 60% of shape errors from our method are under 1.7 mm. The shape error is computed by measuring the nearest point distance from deformed template to raw scan. Overall, the proposed method outperforms all others.

# References

- [1] Alder hey headspace project. <http://www.alderhey.nhs.uk/departments/craniofacial/headspace-project/>.
- [2] Ayman Abaza, Arun Ross, Christina Hebert, Mary Ann F Harrison, and Mark S Nixon. A survey on ear biometrics. *ACM computing surveys (CSUR)*, 45(2):22, 2013.
- [3] Thomas Albrecht, Reinhard Knothe, and Thomas Vetter. Modeling the remaining flexibility of partially fixed statistical shape models. In *2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 160–169, 2008.
- [4] Brett Allen, Brian Curless, and Zoran Popović. Articulated body deformation from range scan data. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 612–619. ACM, 2002.
- [5] Brett Allen, Brian Curless, and Zoran Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM transactions on graphics (TOG)*, volume 22, pages 587–594. ACM, 2003.
- [6] Brett Allen, Brian Curless, and Zoran Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM transactions on graphics*, 22(3):587–594, 2003.
- [7] Brian Amberg, Reinhard Knothe, and Thomas Vetter. Expression invariant 3d face recognition with a morphable model. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

- [8] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid ICP algorithms for surface registration. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007.
- [9] K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700, 1987.
- [10] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. Robust discriminative response map fitting with constrained local models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3444–3451, 2013.
- [11] Timur Bagautdinov, Chenglei Wu, Jason Saragih, Pascal Fua, and Yaser Sheikh. Modeling facial geometry using compositional vaes. *In practice*, 1:1, 2018.
- [12] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 3d constrained local model for rigid and non-rigid facial tracking. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2610–2617. IEEE, 2012.
- [13] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 354–361, 2013.
- [14] Anil Bas and William AP Smith. What does 2d geometric information really tell us about 3d face shape? *arXiv preprint arXiv:1708.06703*, 2017.
- [15] Curzio Basso, Alessandro Verri, and Jens Herder. Fitting 3d morphable models using implicit representations. *Journal of Virtual Reality and Broadcasting*, 4(18):1–10, 2007.
- [16] Michaela Benz. The symmetry of faces. *Proc. of VMV'2002, Erlangen, Germany*, pages 43–50, 2002.

- 
- [17] Robert Bergevin, Denis Laurendeau, and Denis Poussart. Estimating the 3d rigid transformation between two range views of a complex object. In *Pattern Recognition, 1992. Vol. I. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on*, pages 478–482. IEEE, 1992.
- [18] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586–607. International Society for Optics and Photonics, 1992.
- [19] Charles Beumier and Marc Acheroy. Face verification from 3d and grey level clues. *Pattern recognition letters*, 22(12):1321–1329, 2001.
- [20] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, 1999.
- [21] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9):1063–1074, 2003.
- [22] Timo Bolkart and Stefanie Wuhrer. Statistical analysis of 3d faces in motion. In *2013 International Conference on 3D Vision-3DV 2013*, pages 103–110. IEEE, 2013.
- [23] Timo Bolkart and Stefanie Wuhrer. A groupwise multilinear correspondence optimization for 3d faces. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3604–3612, 2015.
- [24] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.
- [25] Fred L Bookstein. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical image analysis*, 1(3):225–243, 1997.

- [26] James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. Large scale 3d morphable models. *International Journal of Computer Vision*, 126(2-4):233–254, 2018.
- [27] James Booth, Anastasios Roussos, Stefanos Zafeiriou, Allan Ponniah, and David Dunaway. A 3d morphable model learnt from 10,000 faces. In *Proceedings of CVPR*, pages 5543–5552, 2016.
- [28] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 3189–3197, 2016.
- [29] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences*, 103(5):1168–1172, 2006.
- [30] Alexander M Bronstein, Michael M Bronstein, Ron Kimmel, Mona Mahmoudi, and Guillermo Sapiro. A gromov-hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *International Journal of Computer Vision*, 89(2-3):266–286, 2010.
- [31] Alan Brunton, Jochen Lang, Eric Dubois, and Chang Shu. Wavelet model-based stereo for fast, robust face reconstruction. In *2011 Canadian Conference on Computer and Robot Vision (CRV)*, pages 347–354, 2011.
- [32] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *International Conference on Computer Vision*, 2017.
- [33] H Cantzler. Random sample consensus (ransac). *Institute for Perception, Action and Behaviour, Division of Informatics, University of Edinburgh*, 1981.
- [34] Will Chang and Matthias Zwicker. Automatic registration for articulated shapes. In *Computer Graphics Forum*, volume 27, pages 1459–1468, 2008.

- 
- [35] Jiansheng Chen, Chang Yang, Yu Deng, Gang Zhang, and Guangda Su. Exploring facial asymmetry using optical flow. *IEEE Signal Processing Letters*, 21(7):792–795, 2014.
- [36] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.
- [37] Peter Claes, Denise K Liberton, Katleen Daniels, Kerri Matthes Rosana, Ellen E Quillen, Laurel N Pearson, Brian McEvoy, Marc Bauchet, Arslan A Zaidi, Wei Yao, et al. Modeling 3d facial shape from dna. *PLoS genetics*, 10(3):e1004224, 2014.
- [38] Peter Claes, Mark Walters, Dirk Vandermeulen, and John Gerald Clement. Spatially-dense 3d facial asymmetry assessment in both typical and disordered growth. *Journal of anatomy*, 219(4):444–455, 2011.
- [39] Thomas F Coleman and Yuying Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6(2):418–445, 1996.
- [40] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):681–685, 2001.
- [41] Timothy F Cootes and Christopher J Taylor. Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing*, 13(5):403–409, 1995.
- [42] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [43] C Creusot, N. E. Pears, and J. Austin. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *Int. Journ. Computer Vision*, 102(1):146–179, 2013.

- [44] Clement Creusot, Nick Pears, and Jim Austin. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International journal of computer vision*, 102(1-3):146–179, 2013.
- [45] David Cristinacce and Timothy F Cootes. Feature detection and tracking with constrained local models. In *Bmvc*, volume 1, page 3. Citeseer, 2006.
- [46] Hang Dai, Nick Pears, and William Smith. Non-rigid 3d shape registration using an adaptive template. In *Proceedings of the European Conference on Computer Vision Workshop Munich, Germany*, 2018.
- [47] Hang Dai, Nick Pears, William Smith, and Christian Duncan. A 3d morphable model of craniofacial shape and texture variation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3104–3112. IEEE, 2017.
- [48] Hang Dai, Nick Pears, William Smith, and Christian Duncan. Symmetric shape morphing for 3d face and head modelling. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 91–97. IEEE, 2018.
- [49] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [50] Rhodri H Davies, Carole J Twining, Tim F Cootes, John C Waterton, and Camillo J Taylor. A minimum description length approach to statistical shape modeling. *Medical Imaging, IEEE Transactions on*, 21(5):525–537, 2002.
- [51] E De Momi, J Chapuis, I Pappas, G Ferrigno, W Hallermann, A Schramm, and M Caversaccio. Automatic extraction of the mid-facial plane for cranio-maxillofacial surgery planning. *International journal of oral and maxillofacial surgery*, 35(7):636–642, 2006.
- [52] Michaël De Smet and Luc Van Gool. Optimal regions for linear model-based 3d face reconstruction. In *Asian Conference on Computer Vision*, pages 276–289, 2010.

- [53] J. Deng, Y. Zhou, S. Cheng, and S. Zafeiriou. Cascade multi-view hourglass model for robust 3d face alignment. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG'18)*, Xi'an, China, May 2018.
- [54] Jiankang Deng, Yuxiang Zhou, Shiyang Cheng, and Stefanos Zaferiou. Cascade multi-view hourglass model for robust 3d face alignment. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 399–403. IEEE, 2018.
- [55] Mathieu Desbrun, Mark Meyer, Peter Schröder, and Alan H Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 317–324. ACM Press/Addison-Wesley Publishing Co., 1999.
- [56] Ian L Dryden, Kanti V Mardia, et al. Statistical shape analysis, 1998.
- [57] Christian Duncan, Hang Dai, Nick Pears, and William Smith. A novel shape based outcomes analysis tool for craniofacial surgeons. In *17th Biennial Congress of the International Society of Craniofacial Surgery (ISCFS)*, 2017.
- [58] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. *Emotion in the Human Face: Guide-lines for Research and an Integration of Findings: Guidelines for Research and an Integration of Findings*. Pergamon, 1972.
- [59] Žiga Emeršič, Dejan Štepec, Vitomir Štruc, Peter Peer, Anjith George, Adii Ahmad, Elshibani Omar, Terranee E Boulton, Reza Safdaii, Yuxiang Zhou, et al. The unconstrained ear recognition challenge. In *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pages 715–724. IEEE, 2017.
- [60] Žiga Emeršič, Vitomir Štruc, and Peter Peer. Ear recognition: More than a survey. *Neurocomputing*, 255:26–39, 2017.
- [61] Anders Ericsson and Johan Karlsson. Measures for benchmarking of automatic correspondence algorithms. *Journal of Mathematical Imaging and Vision*, 28(3):225–241, 2007.

- [62] Timothy C Faltemier, Kevin W Bowyer, and Patrick J Flynn. Using a multi-instance enrollment representation to improve 3d face recognition. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6. IEEE, 2007.
- [63] Olivier D Faugeras and Martial Hebert. The representation, recognition, and locating of 3-d objects. *The international journal of robotics research*, 5(3):27–52, 1986.
- [64] J Michael Fitzpatrick, Jay B West, and Calvin R Maurer. Predicting error in rigid-body point-based registration. *IEEE transactions on medical imaging*, 17(5):694–702, 1998.
- [65] Song Ge, Guoliang Fan, and Meng Ding. Non-rigid point set registration with global-local topology preservation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 245–251. IEEE, 2014.
- [66] Kyle Genova, Forrester Cole, Aaron Maschinot, Aaron Sarna, Daniel Vlasic, and William T Freeman. Unsupervised training for 3d morphable model regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8377–8386, 2018.
- [67] Thomas Gerig, Andreas Forster, Clemens Blumer, Bernhard Egger, Marcel Lüthi, Sandro Schönborn, and Thomas Vetter. Morphable face models - an open framework. *CoRR*, abs/1709.08398, 2017.
- [68] A Golovinskiy, J Podolak, and T Funkhouser. Symmetry-aware mesh processing. *Mathematics of Surfaces XIII*, pages 170–188, 2009.
- [69] Aleksey Golovinskiy, Wojciech Matusik, Hanspeter Pfister, Szymon Rusinkiewicz, and Thomas Funkhouser. A statistical model for synthesis of detailed facial geometry. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 1025–1034, 2006.
- [70] Vladislav Golyanik, Bertram Taetz, Gerd Reis, and Didier Stricker. Extended coherent point drift algorithm with correspondence priors and optimal subsampling.

- 
- In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–9. IEEE, 2016.
- [71] Colin Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 285–339, 1991.
- [72] John C Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.
- [73] Xuan Guo. Three dimensional moment invariants under rigid transformation. In *International Conference on Computer Analysis of Images and Patterns*, pages 518–522. Springer, 1993.
- [74] Daniel Haase, Erik Rodner, and Joachim Denzler. Instance-weighted transfer learning of active appearance models. In *Proceedings of CVPR*, pages 1426–1433, 2014.
- [75] Robin J Hennessy, Stephen McLearie, Anthony Kinsella, and John L Waddington. Facial shape and asymmetry by three-dimensional laser surface scanning covary with cognition in a sexually dimorphic manner. *The Journal of neuropsychiatry and clinical neurosciences*, 18(1):73–80, 2006.
- [76] Thomas Heseltine, Nick Pears, and Jim Austin. Three-dimensional face recognition using combinations of surface feature map subspace components. *Image and Vision Computing*, 26(3):382–396, 2008.
- [77] Michael Hinze and Stefan Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In *Dimension reduction of large-scale systems*, pages 261–306. Springer, 2005.
- [78] Yipeng Hu, Erik-Jan Rijkhorst, Richard Manber, David Hawkes, and Dean Barratt. Deformable vessel-based registration using landmark-guided coherent point drift. In *International Workshop on Medical Imaging and Virtual Reality*, pages 60–69. Springer, 2010.
- [79] IEEE. *A 3D Face Model for Pose and Illumination Invariant Face Recognition*, Genova, Italy, 2009.

- [80] Varun Jain and Hao Zhang. Robust 3d shape correspondence in the spectral domain. In *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)*, pages 19–19. IEEE, 2006.
- [81] Benedict C Jones, Anthony C Little, Ian S Penton-Voak, BP Tiddeman, D Michael Burt, and DI Perrett. Facial symmetry and judgements of apparent health: support for a good genes explanation of the attractiveness–symmetry relationship. *Evolution and human behavior*, 22(6):417–429, 2001.
- [82] Ioannis A Kakadiaris, Dimitri Metaxas, and Ruzena Bajcsy. Active part-decomposition, shape, and motion estimation of articulated objects: A physics-based approach. In *CVPR*, volume 94, pages 980–984, 1994.
- [83] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [84] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2004.
- [85] David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.
- [86] David G Kendall. A survey of the statistical theory of shape. *Statistical Science*, pages 87–99, 1989.
- [87] John T Kent and Kanti V Mardia. Shape, procrustes tangent projections and bilateral symmetry. *Biometrika*, pages 469–485, 2001.
- [88] Hyeonwoo Kim, Michael Zollhöfer, Ayush Tewari, Justus Thies, Christian Richardt, and Christian Theobalt. Inversefacenet: Deep single-shot inverse face rendering from a single image. *arXiv preprint arXiv:1703.10956*, 2017.
- [89] Christian Peter Klingenberg, Marta Barluenga, and Axel Meyer. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution*, 56(10):1909–1920, 2002.

- [90] R. R. Kolamunnage-Dona and J. T. Kent. Decomposing departures from bilateral symmetry. In S. Barber, P.D. Baxter, K. V. Mardia, and R. E. Walls, editors, *Quantitative Biology, Shape Analysis, and Wavelets*, pages 75–78. Leeds University Press, 2005.
- [91] Aaron CW Kotcheff and Chris J Taylor. Automatic construction of eigenshape models by direct optimization. *Medical Image Analysis*, 2(4):303–314, 1998.
- [92] Hao Li, Robert W. Sumner, and Mark Pauly. Global correspondence optimization for non-rigid registration of depth scans. *Eurographics Symposium on Geometry Processing*, 27(5):1421–1430, 2008.
- [93] Yanxi Liu and Jeff Palmer. A quantified study of facial asymmetry in 3d faces. 2003.
- [94] Marcel Lüthi, Thomas Gerig, Christoph Jud, and Thomas Vetter. Gaussian process morphable models. *IEEE transactions on pattern analysis and machine intelligence*, 2017.
- [95] Kanti V Mardia, Fred L Bookstein, and Ian J Moreton. Statistical assessment of bilateral symmetry of shapes. *Biometrika*, pages 285–300, 2000.
- [96] Iain Matthews and Simon Baker. Active appearance models revisited. *International journal of computer vision*, 60(2):135–164, 2004.
- [97] David Milner, Shmuel Raz, Hagit Hel-Or, Daniel Keren, and Eviatar Nevo. A new measure of symmetry and its application to classification of bifurcating structures. *Pattern Recognition*, 40(8):2237–2250, 2007.
- [98] Niloy J Mitra, Leonidas J Guibas, and Mark Pauly. Partial and approximate symmetry detection for 3d geometry. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 560–568. ACM, 2006.
- [99] Ana Belen Moreno and Armand Sánchez. Gavabdb: a 3d face database. In *Proc. 2nd COST275 Workshop on Biometrics on the Internet, Vigo (Spain)*, pages 75–80, 2004.

- [100] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010.
- [101] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12):2262–2275, 2010.
- [102] Maks Ovsjanikov, Quentin Mérigot, Facundo Mémoli, and Leonidas Guibas. One point isometric matching with the heat kernel. In *Computer Graphics Forum*, volume 29, pages 1555–1564, 2010.
- [103] Maks Ovsjanikov, Jian Sun, and Leonidas Guibas. Global intrinsic symmetries of shapes. In *Computer graphics forum*, volume 27, pages 1341–1348. Wiley Online Library, 2008.
- [104] Viorica Patraucean, Rafael Grompone von Gioi, and Maks Ovsjanikov. Detection of mirror-symmetric image patches. In *Proc. CVPR Workshops*, pages 211–216, 2013.
- [105] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *Advanced video and signal based surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 296–301, 2009.
- [106] Nick Pears and C V Jan. Automatic 3D modelling of craniofacial form. *arXiv*, pages 1–57, 2016.
- [107] Yuri Pekelnny and Craig Gotsman. Articulated object reconstruction and markerless motion capture from depth video. In *Computer Graphics Forum*, volume 27, pages 399–408, 2008.
- [108] Anika Pflug and Christoph Busch. Ear biometrics: a survey of detection, feature extraction and recognition methods. *IET biometrics*, 1(2):114–129, 2012.
- [109] P Jonathon Phillips, Patrick J Flynn, Todd Scruggs, Kevin W Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the

- face recognition grand challenge. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 947–954. IEEE, 2005.
- [110] Joshua Podolak, Philip Shilane, Aleksey Golovinskiy, Szymon Rusinkiewicz, and Thomas Funkhouser. A planar-reflective symmetry transform for 3d shapes. *ACM Transactions on Graphics (TOG)*, 25(3):549–559, 2006.
- [111] Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. Generating 3d faces using convolutional mesh autoencoders. *arXiv preprint arXiv:1807.10267*, 2018.
- [112] B Robertson, H Dai, N Pears, and C Duncan. A morphable model of the human head validating the outcomes of an age-dependent scaphocephaly correction. *International Journal of Oral and Maxillofacial Surgery*, 46:68, 2017.
- [113] Augusto Salazar, Stefanie Wuhrer, Chang Shu, and Flavio Prieto. Fully automatic expression-invariant face correspondence. *Machine Vision and Applications*, 25(4):859–879, 2014.
- [114] Arman Savran, Neşe Alyüz, Hamdi Dibeklioglu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Bosphorus database for 3d face analysis. In *European Workshop on Biometrics and Identity Management*, pages 47–56. Springer, 2008.
- [115] Yoland Savriama and Christian Peter Klingenberg. Beyond bilateral symmetry: geometric morphometric methods for any type of symmetry. *BMC evolutionary biology*, 11(1):280, 2011.
- [116] Ryan Schmidt and Karan Singh. Meshmixer: an interface for rapid mesh composition. In *ACM SIGGRAPH 2010 Talks*, page 6. ACM, 2010.
- [117] Muhammad Sharif, Kamran Ayub, Danish Sattar, Mudassar Raza, and Sajjad Mohsin. Enhanced and fast face recognition by hashing algorithm. *Journal of applied research and technology*, 10(4):607–617, 2012.

- [118] Aurela Shehu, Alan Brunton, Stefanie Wuhrer, and Michael Wand. Characterization of partial intrinsic symmetries. In *European Conference on Computer Vision*, pages 267–282. Springer, 2014.
- [119] Edgar Simo-Serra, Eduard Trulls, Luis Ferraz, Iasonas Kokkinos, Pascal Fua, and Francesc Moreno-Noguer. Discriminative learning of deep convolutional feature point descriptors. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 118–126, 2015.
- [120] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pages 175–184, 2004.
- [121] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, pages 109–116, 2007.
- [122] Olga Sorkine, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and H-P Seidel. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 175–184. ACM, 2004.
- [123] Martin A Styner, Kumar T Rajamani, Lutz-Peter Nolte, Gabriel Zsemlye, Gábor Székely, Christopher J Taylor, and Rhodri H Davies. Evaluation of 3d correspondence methods for model building. *Information processing in medical imaging*, pages 63–75, 2003.
- [124] Gary KL Tam, Zhi-Quan Cheng, Yu-Kun Lai, Frank C Langbein, Yonghuai Liu, David Marshall, Ralph R Martin, Xian-Fang Sun, and Paul L Rosin. Registration of 3d point clouds and meshes: a survey from rigid to nonrigid. *IEEE transactions on visualization and computer graphics*, 19(7):1199–1217, 2013.
- [125] Qingyang Tan, Lin Gao, Yu-Kun Lai, and Shihong Xia. Variational autoencoders for deforming 3d mesh models. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

- [126] Jos MF Ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977.
- [127] Frank B ter Haar and Remco C Veltkamp. 3d face model fitting for recognition. In *European Conference on Computer Vision*, pages 652–664, 2008.
- [128] Demetri Terzopoulos and Dimitri Metaxas. Dynamic 3d models with local and global deformations: Deformable superquadrics. In *Computer Vision, 1990. Proceedings, Third International Conference on*, pages 606–615. IEEE, 1990.
- [129] Ayush Tewari, Michael Zollhöfer, Pablo Garrido, Florian Bernard, Hyeonwoo Kim, Patrick Pérez, and Christian Theobalt. Self-supervised multi-level face model learning for monocular reconstruction at over 250 hz. *arXiv preprint arXiv:1712.02859*, 2, 2017.
- [130] Darcy Wentworth Thompson et al. On growth and form. *On growth and form.*, 1942.
- [131] Luan Tran and Xiaoming Liu. Nonlinear 3d face morphable model. *arXiv preprint arXiv:1804.03786*, 2018.
- [132] Philip A Tresadern, Patrick Sauer, and Timothy F Cootes. Additive update predictors in active appearance models. In *Proceedings of BMVC*, volume 2, page 4, 2010.
- [133] Imen Hamrouni Trimech, Ahmed Maalej, and Najoua Essoukri Ben Amara. 3d facial expression recognition using nonrigid cpd registration method. In *Information and Digital Technologies (IDT), 2017 International Conference on*, pages 478–481. IEEE, 2017.
- [134] Laurens Van Der Maaten. Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1):3221–3245, 2014.
- [135] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

- [136] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. Face transfer with multilinear models. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 426–433, 2005.
- [137] Peng Wang, Ping Wang, ZhiGuo Qu, YingHui Gao, and ZhenKang Shen. A refined coherent point drift (cpd) algorithm for point set registration. *Science China Information Sciences*, 54(12):2639–2646, 2011.
- [138] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [139] Zhou Wang and Qiang Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, 2011.
- [140] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.
- [141] Max Wardetzky, Miklós Bergou, David Harmon, Denis Zorin, and Eitan Grinspun. Discrete quadratic curvature energies. *Computer Aided Geometric Design*, 24(8-9):499–518, 2007.
- [142] Lingyu Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1544–1553, 2016.
- [143] Yijun Wu, Gang Pan, and Zhaohui Wu. Face authentication based on multiple profiles extracted from range data. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 515–522, 2003.
- [144] Yue Wu, Tal Hassner, KangGeon Kim, Gerard Medioni, and Prem Natarajan. Facial landmark detection with tweaked convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 2017.

- 
- [145] Yue Wu and Qiang Ji. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, pages 1–28, 2017.
- [146] Fei Yang, Lubomir Bourdev, Eli Shechtman, Jue Wang, and Dimitris Metaxas. Facial expression editing in video using a temporally-smooth factorization. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 861–868. IEEE, 2012.
- [147] Fei Yang, Jue Wang, Eli Shechtman, Lubomir Bourdev, and Dimitri Metaxas. Expression flow for 3d-aware face component transfer. In *ACM Transactions on Graphics (TOG)*, volume 30, page 60, 2011.
- [148] Yi Yang and Deva Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1385–1392. IEEE, 2011.
- [149] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3d dynamic facial expression database. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.
- [150] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J Rosato. A 3d facial expression database for facial behavior research. In *7th international conference on automatic face and gesture recognition (FGR06)*, pages 211–216. IEEE, 2006.
- [151] Rui Yu, Chris Russell, Neill Campbell, and Lourdes Agapito. Direct, dense, and deformable: Template-based non-rigid 3d reconstruction from rgb video. In *IEEE International Conference on Computer Vision (ICCV 2016)*, 2015.
- [152] Kun Zhang, Yuan Cheng, and Wee Kheng Leow. Dense correspondence of skull models by automatic detection of anatomical landmarks. In *Computer Analysis of Images and Patterns*, pages 229–236, 2013.
- [153] Liyan Zhang, Anshuman Razdan, Gerald Farin, John Femiani, Myungsoo Bae, and Charles Lockwood. 3d face authentication and recognition based on bilateral symmetry analysis. *The Visual Computer*, 22(1):43–55, 2006.

- [154] Cheng Zhong, Zhenan Sun, and Tieniu Tan. Robust 3d face recognition using learned visual codebook. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007.
- [155] Yuxiang Zhou and Stefanos Zafeiriou. Deformable models of ears in-the-wild for alignment and recognition. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 626–633. IEEE, 2017.
- [156] Zhiyong Zhou, Jian Zheng, Yakang Dai, Zhe Zhou, and Shi Chen. Robust non-rigid point set registration using student’s-t mixture model. *PloS one*, 9(3):e91381, 2014.
- [157] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of CVPR*, pages 2879–2886, 2012.
- [158] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. Face alignment across large poses: A 3d solution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 146–155, 2016.
- [159] Alexei I Zhurov, Stephen Richmond, Chung How Kau, and Arshed Toma. Averaging facial images. *Three-dimensional imaging for orthodontics and maxillofacial surgery*. London: Wiley-Blackwell, pages 126–44, 2010.
- [160] Reza Zolfaghari, Nicolas Epain, Craig T Jin, Joan Glaunès, and Anthony Tew. Generating a morphable model of ears. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 1771–1775. IEEE, 2016.