Automated Sex Identification from 3D Skulls based on Patch Representations for Forensic Examinations



by

Olasimbo Ayodeji Arigbabu

School of Computer Science

A thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

April 7, 2019

Abstract

This thesis introduces new perspective of addressing skull sex determination by investigating the possibility of using computer vision based solutions for sex determination from skull data, and developing new algorithms tailored for mitigating the underlying problems experienced in the conventional forensic sex determination methods.

First, this thesis proposes the use of state-of-the-art 3D local shape descriptors to derive patch based representations from the skull. This is achieved by geometrically partitioning the skull into local regions, and aggregating the features from the regions into compact representation, which summarize the regional shape properties. Several rigorous experiments are conducted using this approach, and the results show that the proposed patch based representation is comparable to conventional forensic methods with a sex prediction rate of 86%. Traditional morphometric analysis on the same dataset produced prediction rate of 86.4%, while PCA/LDA based method produced prediction rate of 75% which indicates that the proposed patch based representation is competitive and an effective method of sex determination.

Second, this thesis proposes a new approach of partitioning the skull into pseudo-anatomical regions, which is reminiscent of the anatomical regions commonly examined in forensic morphological assessment. In order to achieve this, the skull samples are registered to establish correspondence across the dataset using an improved version of coherent point drift (CPD) algorithm named Ensemble CPD. The correspondence established is exploited to partition the skulls using Fuzzy cmean (FCM) into regions which share visual resemblance with the anatomical regions in forensic morphological assessment. Further, 3D local shape descriptors are computed from each region, aggregated into compact representation, and eventually used for sex determination. The best result in this experimental evaluation is 86.5% which is also comparable to conventional forensic methods.

Finally, this thesis proposes two group variable selection algorithms which are (1) novel group regression model for manifold data and (2) novel group regression model for manifold data with split augmented lagrangian algorithm, for automatically discovering discriminant (important) skull regions. The main objective of these methods is to prune out the salient regions of the skull that contribute positively to sex prediction performance. Forensic methods have demonstrated that some anatomical parts are more sexually dimorphic than others, however such analysis is performed using stepwise variable selection. Thus, this thesis introduces new techniques for automatically learning such information directly from the data. The best result in this experiment is 84.5% which is comparable to conventional forensic methods.

Acknowledgements

First and foremost, I thank Almighty Allah (Subhanahu Wa Ta'ala) for his mercy, guidance, and protection during the journey of my doctoral program and I send blessings upon prophet Muhammad (Sallallahu 'Alaihi Wa Salam).

I would like to extend my sincere gratitude to my supervisor, Assc. Prof Iman Yi Liao, for supporting me with her knowledge, wisdom, experience, and dedication. For initiating several intuitive conversations that immensely improve the outcome of this research work. Also, for helping me secure financial support from University of Nottingham and the Ministry of Science, Technology and Innovation (MOSTI) which made all the work in this thesis possible.

I thank my co-supervisors Assc. Prof Chong Siang Yew, Assc. Prof Bai Li, and Dr. Nurliza Abdullah for providing me with very helpful suggestions and reading the reports emanating from this research.

The continued moral, emotional, and spiritual support of my family was invaluable throughout. Thank you very much for being there at the most important period of my studies. I am very grateful to you.

Contents

1 Introduction **1**

1.1 Introduction 1
1.2 Background 2
1.3 Motivation 3
1.4 Research Questions 4
1.5 Objectives and Scope 5
1.6 Thesis Contributions 6
1.7 Publications 7

2 Literature Review **9**

2.1 Cranial Sex Estimation in Forensic Anthropology 9
2.1.1 Approaches based on Morphological Assessment 10
2.1.2 Approaches based on Morphometric Analysis 14
2.2 Gender Recognition in Computer Vision 19
2.2.1 Geometric based 20
2.2.2 Appearance based 22
2.3 Relation between Forensic and Computer Vision Methods 25
2.4 3D Local Descriptors and Patch based Representation 26
2.4.1 3D Shape Representation for Object Recognition 26
2.4.2 3D Shape Representation for Craniofacial Analysis 30
2.5 Conclusion 31

3 Research Methodology **32**

3.1 Overview 32
3.1.1 Experimental Dataset 33
3.2 3D Preprocessing 35
3.2.1 Noise Removal 35
3.2.2 Background Object Removal 37
3.2.3 Mesh Alignment 39
3.3 3D Local Shape Descriptors 40

3.3.1 Mesh Local Binary Pattern (meshLBP) . . . 41
3.3.2 Spin Image . . . 41
3.3.3 Local Depth SIFT (LD-SIFT) . . . 43
3.3.4 Scale Invariant Heat Kernel Signature (SIHKS) . . . 43
3.3.5 ShapeDNA . . . 44
3.3.6 Area of Skull Regions . . . 45
3.4 Feature Learning . . . 45
3.4.1 KPCA . . . 45
3.4.2 K-SVD . . . 46
3.5 Classification . . . 46
3.5.1 Support Vector Predictive Models . . . 47
3.5.2 Kernel Extreme Learning Machine (K-ELM) . . . 47
3.5.3 Sparse Representation Classifier . . . 48
3.5.4 Multiple Kernel Learning (MKL) . . . 49

4 An Empirical Study on the Typical 3D Shape Descriptors and their Patch Representations for Sex Identification 51
4.1 Overview . . . 51
4.2 Patch based Representation and Feature Aggregation from Heuristically Partitioned Skulls . . . 52
4.3 Sex Determination using Patch based Representations from Heuristically Partitioned Skulls . . . 54
4.3.1 Initial Analysis using BoW representation . . . 55
4.3.2 Initial Analysis using Stacked Representation . . . 55
4.3.3 Effect of Number of Patches and Partition Orientation on Sex Prediction . . . 59
4.3.4 Sex determination using MKL on Multiple Features . . . 62
4.4 Conclusion . . . 69

5 A Pseudo Anatomical Region based Patch Representation 71
5.1 Overview . . . 71
5.2 Related Works on Skull Partitioning . . . 72
5.3 Pseudo-anatomical Partitioning . . . 74
5.3.1 Coherent Point Drift . . . 74
5.3.2 Ensemble CPD . . . 75
5.3.3 Application of Ensemble CPD to Skull Registration and Partitioning . . . 80
5.3.4 Feature Extraction on Skull Partitions . . . 80

5.4 Sex Determination from Pseudo-anatomical Patches 83
5.4.1 Effect of Number of Pseudo-anatomical Regions on Sex Determination 83
5.4.2 Sex Determination using MKL 88
5.5 Conclusion 89

6 Novel Methods for Selecting Discriminant Anatomical Regions 91
6.1 Overview 91
6.2 Related Works on Skull Region Selection on Sex Identification . . . 92
6.2.1 Approaches to Skull Region Selection in Forensic Anthropology 93
6.2.2 Group Variable Selection 94
6.3 Region Selection 96
6.3.1 Group Lasso 96
6.3.2 Novel Group Regression Model for Manifold Data 98
6.3.3 Novel Group Regression Model for Manifold Data with Split Augmented Lagrangian Shrinkage Algorithm (SALSA) . . . 101
6.3.4 Multi-scale Region Learning/Selection 104
6.4 Sex Determination 105
6.5 Conclusion 110

7 Validation Study 112
7.1 Effect of Data Splitting on Sex Determination 112
7.2 Effect of Population Group on Sex Determination 113
7.3 Comparison with Forensic Approach 114
7.3.1 Morphometric Analysis 114
7.3.2 Statistical Analysis 115
7.3.3 Patch Representation and Keypoints Analysis 116

8 Discussion and Conclusion 118
8.1 Future Works 120

A Prototype 122

Bibliography 125

List of Figures

2.1 Standardized scale and diagram for scoring cranial traits (Buikstra and Ubelaker., 1994) 11
2.2 Standard landmarks of sexually dimorphic traits of the cranium (Franklin et al., 2012) 16
2.3 Face measurements performed by Fellous (1997) 21
2.4 Extraction of primary features of the face (Han et al., 2009) 22

3.1 Structure of sex determination using heuristically partitioned skulls 32
3.2 Distribution of the 100 data samples used in the experiments in this thesis according to sex, age, and race 33
3.3 Head region volume CT slices of 4 subjects prior to 3D reconstruction 34
3.4 3D reconstruction from volume slices using Marching cubes algorithm 37
3.5 Examples of Reconstructed Skull data. (a) Original Skull (b) Downsampled to 50% (b) Downsampled to 25% (d) Downsampled to 13%. 37
3.6 Noisy objects removal (a) data with noise (b) clean and aligned skulls after object removal (c) the removed noise 40
3.7 Example of MeshLBP feature distribution by mapping the binary patterns to different color map. For illustration, the binary patterns of ring (r = 2) are displayed. Red color indicates areas of small binary pattern (0) due to small variation in the curvature of the local regions. Blue represents patterns with values (1 16), green denotes patterns with values (32 256), and cyan represents patterns with values (512 4096), due to large variation in the curvature of the local regions. For instance, sharp curvature variation can be observed around the local regions of the orbital, nasal bone, and mandible, while the parietal bone exhibits mild variation in curvature 42
3.8 (a) and (c) Eigenvector over the rendered 3D skull surface. The areas with small eigenvectors are represented with color blue and as the values increase, the color goes from green to red (b) and (d) Plot of 100 eigenvalues forming ShapeDNA descriptor. 45

4.1 Structure of sex determination using heuristically partitioned skulls 52
4.2 An example of multi-region representation with LD-SIFT. Each color denotes a partitioned sub-region of the skull. 53
4.3 Illustration of multi-region feature representations using different number of partitions. (b)-(d) show skull partitioning performed along the Z-axis on skull sample 1 54
4.4 Some examples of multi-region partitioned skulls along different axes. (a) on skull sample 7 (b) on skull sample 4 54
4.5 **A valid SRC test.** The skull on the left belongs to subject number 11, which is in the test set. The figure on the right illustrates the learned coefficients on the training matrix, with largest coefficient found as subject number 8. As a result, the test shows that these two samples belong to the same class (male). 57
4.6 Student's T-test conducted on extracted features (a) Long tailed feature representation (b) Stacked feature representation 58
4.7 Results of sex estimation based on orientation of region partition and number of regions. 60
4.8 Results of sex estimation based on orientation of region partition and number of regions using KSVD. 63
4.9 Results of sex estimation based on orientation of region partition and number of regions using KPCA. 65
4.10 Results of sex estimation based on orientation of region partition and number of regions using MKL 68
4.11 An example of sex identification using hard margin MKL with primal formulation for 3D feature combination. 69

5.1 Structure of sex determination using anatomically partitioned skulls 72
5.2 Comparison of registration methods. Blue color indicates the source (model) data, while red color is the registered data. 78
5.3 Registration error with respect to iterations on the 5 datasets. . . . 79
5.4 Pseudo-anatomical Skull partitions of 6 different samples showing the locations of 5 regions (patches) of the skulls. 81
5.5 Illustration of patch representation of 3D local shape descriptors from pseudo-anatomical regions. 82
5.6 Sex prediction results for different number of regions (1 - 99) from pseudo-anatomical partitions using different feature extraction and classification methods. 88

5.7 Illustration of patch representation of 3D local shape descriptors from pseudo-anatomical regions. 89

6.1 Structure of skull region selection and sex determination 92
6.2 Illustration of pseudo-anatomical partitioning with FCM 106
6.3 Illustration of group patch selection from the clustered regions by group LASSO 107
6.4 Illustration of group patch selection from the clustered regions using the proposed methods 108

7.1 Figure of prediction rate for each partition. 116

A.1 Screenshot of the designed prototype showing the 3D planes of 3D CT volume 123
A.2 Example of skull reconstruction 123
A.3 Illustration of shape primitive function with an example depicting mean curvature 124
A.4 shows the extracted MeshLBP and the prediction results 124

List of Tables

2.1 3D Shape Descriptors 27

3.1 Number of samples in different age band 34
3.2 Number of samples in race groups 34

4.1 Sex prediction results using long tailed features with BoW representation 55
4.2 Sex prediction Results with stacked features 56

5.1 Computational time of registration 78
5.2 Registration error based on distance between model and registered target 79

6.1 Result of skull region selection based on group variable selection . . 109
6.2 Results of measuring the correlation coefficient of two different skull regions 110

7.1 Performance comparison based on data splitting method 113
7.2 Performance comparison based on population group 114
7.3 Performance comparison with Traditional Morphometrics 115
7.4 Performance comparison with Traditional Morphometrics 116

List of Abbreviations

2-Dimension	2D
3-Dimension	3D
Alternating Direction Method of Multipliers	ADMM
Artificial Neural Network	ANN
Augmented Langragian Method	ALM
Bag of Words	BoW
Coherent Point Drift	CPD
Computed Tomography	CT
Convolutional Neural Network	CNN
Difference of Gaussians	DoG
Discrete Cosine Transform	DCT
Discriminant Function Analysis	DFA
Extreme Learning Machine	ELM
Fuzzy C-Means	FCM
Gaussian Mixture Model	GMM
Generalized Cross Validation	GCV
Generalized Procrustes Analysis	GPA
Heat Kernel Signature	HKS
Iterative Closest Point	ICP
K-Singular Value Decomposition	KSVD
Kernel Principal Component Analysis	KPCA
Least Absolute Shrinkage and Selection Operator	LASSO
Linear Discriminant Analysis	LDA
Local Binary Pattern	LBP
Multi-Kernel Learning	MKL
Non Negative Matrix Factorization	NNMF
Penalized Least Square	PLS
Post Mortem Computed Tomography	PMCT
Radial Basis Function	RBF
Scale Invariant Feature Transform	SIFT
Sparse Representation Classifier	SRC
Split Augmented Lagrangian Shrinkage Algorithm	SALSA
Support Vector Machine	SVM
Thin Plate Spline	TPS

Chapter 1

Introduction

1.1 Introduction

> "We shall not cease from exploration. And the end of all our exploring will be to arrive where we started and know the place for the first time."
>
> Thomas Stearns Eliot

Estimation of demographic information such as sex, age, and ancestry from human skeletal remains, particularly the skull region is a challenging task which has attracted significant amount of interest in the forensic anthropology research community. This form of identification, otherwise regarded as a circumstantial identification is crucial in analysing the biological profile to build a comprehensive statistical distribution of the victims involved in tragic event, or used as supporting information in forensic investigation. Through decades of rigorous scientific research, forensic anthropologists have developed substantial understanding of the anatomical regions that contribute to skull demographic estimation. A forensic expert who is well experienced in the anatomical variation in a particular population can confidently predict the age, sex, and ancestry of an unknown skull. However, the prediction can deteriorate rapidly with high statistical significance when assessing skulls from a different population. Such discrepancies can be attributed to the predominant effect of subjectivity of the human observer on the perceived variation in the traits. Moreover, annotation of anatomical traits and/or sites of the skull requires immense precision and accuracy which is time consuming and necessitates sound domain knowledge.

This thesis introduces a novel conceptual approach where the interest is to investigate the possibility of using robust local feature representations and machine learning predictive models that have gained prominence in computer vision research

domain for sex determination from human skull. The research provides enlightenment on, how and the extent to which, such descriptors and predictive models can be used for sex estimation. Though, local descriptors have been demonstrated to be effective in providing demographic estimates from 2D and 3D facial data, their exploits on 3D skull data which has no textural skin properties and primary facial attributes such as eyes, eyebrow, nose, ear, and mouth is yet to be investigated. Despite the fact that such feature descriptors do not explicitly describe specific anatomical details, as in the case of conventional forensics, they implicitly embed local representations which consider the relative local topological relationship of points in a 2D manifold space (i.e surface). Likewise, few research works in forensic analysis have exhibited the potential of using predictive models to determine the class of a given skull. However, these are only limited to determination of sex class from the pre-allocated skull landmark configurations.

On the one hand, this research analyses how these local descriptors can be automatically extracted, compactly represented, and learned for predicting sex from 3D skull data. On the other hand, it demonstrates how local skull regions, which contribute to the performance of sex determination, can be automatically discovered through learning process. This introductory chapter details a brief general overview of the background studies and the main motivation for embarking on this research. It points out the research questions, objectives of the study and the research contributions. In the subsequent parts of this thesis, the following terms are used interchangeably: sex and gender, because in forensic anthropology the term: sex is typically used in the literature, while computer vision experts refer to it as gender.

1.2 Background

Identification of human skeletal remains after occurrence of natural disaster or tragic event of some sort, is highly important in forensic science and anthropology (Scheuer, 2002). Though, forensic experts would examine the retrieved remains through traditional methods such as DNA analysis and dental examination, a substantial amount of victims still remain unidentifiable. In a number of cases, approximating the possible identity of a person through computer aided concepts such as three dimensional (3D) facial reconstruction or superimposition is usually resorted to, especially when no other identification evidence can be obtained with traditional methods (Aulsebrook et al., 1995; Ibáñez et al., 2015). Despite that, it has been strongly advised that facial reconstruction should be restricted

to excluding identity if gross incompatibilities are present (Cattaneo, 2007). Alternatively, forensic experts may seek to build the statistical background profile (conceptually referred to as osteobiography) of humans by assessing the sustained skeletal remains. From the literature, the highly regarded biological information worth documenting include sex, age, and ancestry (Casado, 2017; Giurazza et al., 2012; Popa et al., 2009; Schmitt et al., 2006). Though these traits can not confirm a positive identification of the skeletal remains, they provide a detailed categorization of the subjects. For instance, when the sex of a skeletal sample is determined with high confidence, half of the entire database can be excluded, which can be further simplified by estimating the age and ancestry. These human demographic characteristics have a long dated history in forensic anthropology, and skeletal regions such as pelvis, limbs, skull, and mandible have been thoroughly investigated in previous studies. In this study, the focus is placed on determining sex from the skull region, because it is the most retained part after body decomposition.

1.3 Motivation

Forensic experts have developed various techniques for estimating sex from human skull. Technically, the estimation process involves morphological assessment and morphometric analysis of skeletal sites. These two approaches concentrate on specific anatomical locations of the skull. For instance, morphological methods provide verbal descriptions of anatomical regions before quantifying the descriptions in an ordinal scale, while morphometrics involves linear or geometric measurements of anatomical landmarks. Both techniques eventually adopt discriminant function analysis to attain the demographic prediction. While these techniques appear very simplistic and only make use of discrete amount of traits, they are quite effective in providing solutions to identification of skeletal remains. Moreover, the concepts have been established on well grounded principles and universally agreed on, to be used as evidence in court cases and forensic investigation. However, it is essential to mention that these techniques possess several limitations. First is the influence of observer subjectivity on the confidence of prediction. Since morphological assessment involves verbal descriptions, it naturally implies that humans would exhibit differing conclusion in their visual perception. This influence has been well reported in previous studies as inter-observer variability, especially when forensic expert make prediction on skull samples they are unfamiliar with (Bigoni et al., 2010; Lewis and Garvin, 2016). Likewise, inter-observer variability is inherent when observers select landmarks in the case of morphometric analysis. Furthermore, this influence of population difference affects the accuracy and precision of

the traits used for identification. As a result, an estimation method used in a specific population may not generalize to other populations. Second is the problem of annotation time required for accurate landmark selection from anatomical regions of the skull. Besides, it often requires expensive, specialized anthropometric equipment. In addition forensic experts usually face the challenge of performing measurements that capture subtle variations among cranial traits that are easy to see but very difficult to measure (Walker, 2008). These aforementioned limitations are the core motivations for this study. It is hoped through automatic feature extraction and learning of compact, discriminant features, the representation gap between humans and cranial traits can be bridged. In essence, the effect of subjectivity in visually assessed traits and incompleteness of anatomical landmarks can be eliminated, as well as minimizing the time required for estimation. To further buttress the points, these limitations are well elaborated in chapter 2 by reviewing the previous related works.

1.4 Research Questions

The following research questions are posed in this thesis:

- **Can local feature representation methods be utilized for sex estimation from skull data?** This question is pertinent to the essence of this research. Unlike facial image (either 2D or 3D), where skin texture and the shape of the eyes, eyebrow, nose, and mouth can distinguish two faces, the skull generally has very little distinctive information, which fundamentally requires the knowledge of a forensic expert to discern. It will be important to understand how 3D local descriptors can capture these subtle information.

- **How can these local descriptors be used to provide meaningful and comprehensive representations?** Due to the complex structural configuration of 3D models, descriptions from different skull samples may share similar feature values, which make them less distinctive for distinguishing between the two classes. Hence, it will be essential to derive techniques for connoting comprehensive yet compact representations of the local feature, which makes them more discriminant.

- **Are certain regions of the skull more important than others?** While finding methods to represent the local skull regions in a discriminant form, it will be interesting to investigate whether the entire skull is needed for sex determination or some local skull regions possess better discriminant capabilities than others.

Finding solution to these questions is a compelling premise to this work, to investigate the possibility of bridging the representation gap between conventional forensic anthropologist methods and cranial traits. Its conformity to forensic standards would make the research more useful in forensic examinations.

1.5 Objectives and Scope

The overall aim of this research is to investigate and develop computational methods for skull sex determination composing of several sub-processes of 3D preprocessing, 3D feature representation, skull region learning, sex class prediction based on machine learning predictive models. This requires conducting an in-depth investigation on how local regions of the skull can be compactly, yet comprehensively represented and learned. Further, it needs deep understanding of which local features contribute positively to the sex determination problem. To achieve this, the following objectives are to be attained:

- To evaluate the state-of-the-art patch based 3D shape descriptors on sex identification of 3D skulls. The use of local descriptors extracted on 3D objects have been demonstrated in the aspects of face and object recognition. However, there are no studies reporting the potential of these descriptors on 3D skull, which possesses no textural properties unlike face and arbitrary 3D objects. Thus, the first objective of this research is to demonstrate this possibility by forming patch based representations of the skull with 3D descriptors for sex identification and performing comparative assessment of the representations derived from different patches.

- To propose alternative patch based representations that are reminiscent of anatomical regions on skulls and their application to sex identification of 3D skulls. The second objective of this research is concerned with investigating the possibility of forming representations which take into account the dynamism of the geometric structure of the skull. The aim is to propose computational approach for representing the anatomical shape of the skull as patches with 3D local descriptors. Under this approach, is the introduction of an automatic technique of partitioning the skull anatomically, which serves as an alternative to the manually segmented skull regions under conventional morphological assessment.

- To propose a feature selection method that can select patches (regions) reminiscent of important anatomical structures for sex identification of 3D skulls.

The third objective of this thesis is to introduce the possibility of automatically determining the crucial skull regions which influence prediction performance based on the concept of group variable selection. This follows the convention demonstrated in standard forensic analysis that combination of a set of skull regions are usually enough for prediction and the entire skull may not be necessary for prediction.

- To evaluate and compare the proposed methods with traditional techniques in forensic anthropology. In order to justify the proposed methods, the final objective of this thesis is to compare the performance with traditional morphometric analysis and statistical analysis based approaches.

1.6 Thesis Contributions

The focus of thesis is to investigate the suitability of computer vision based solutions for sex identification, which offer a different perspective of mitigating the subjectivity experienced when morphological assessments are used, and to provide an alternative to morphometrics, which naturally requires a lot of computational or human effort to find very precise landmark points. Also, to investigate how important local skull regions can be automatically determined based on their impact on identification results. As a result, the main contributions of the thesis can be summarized as follows:

- The thesis completed a thorough analysis on a number of popular computer vision based solutions on skull data and confirmed that it is possible to employ such solutions with refinement to perform sex identification from 3D skull data. The analysis is performed by using 3D local descriptors to form patch based representation of the skull obtained through geometric partitioning.

- A new approach for partitioning the skull into pseudo-anatomical regions based on fuzzy cmeans (FCM) is introduced, which is reminiscent of the anatomical regions in conventional forensic morphological assessment. Under this method, a registration method named Ensemble coherent point drift (Ensemble CPD) is proposed which improves the standard CPD algorithm, most especially in terms of computational cost and time. The algorithm is used to establish correspondence across the entire dataset and aid propagation of the locations of partitioned skull regions across the entire dataset. Eventually the pseudo-anatomical regions are used for sex determination of the 3D skulls.

- Two novel algorithms: (1) Novel Group Regression Model for Manifold Data and (2) Novel Group Regression Model for Manifold Data with Split Augmented Lagrangian Algorithm, which are based on the concept of group variable selection are proposed for automatically learning and discovering discriminant skull regions. The aim of these techniques is to prune out the discriminant subset of features that contribute positively to sex prediction problem from large set of features extracted from the skull. The algorithms improve on the standard group least absolute shrinkage and selection operator (LASSO) in terms of consistency of the selected skull regions and classification accuracy.

Methods that are similar to the proposed computer vision based solutions for sex identification are quite limited in the literature. A thorough search of relevant literature provided only one related work, which is the technique proposed by Luo et al. (2013a). However, the method mainly addressed the skull globally using statistical representation of the skull without focusing on the local details. Moreover, the 3D skull data used in the analysis is population specific.

1.7 Publications

The following are the papers resulting from the research presented in this thesis, which have been published, submitted, or under review:

- Arigbabu O.A., Iman Y.L., Abdullah N., Helmee M. Can Computer Vision Techniques be Applied to Automated Forensic Examinations? A Study on Sex Identification from Human Skulls Using Head CT Scans. In: Lai SH., Lepetit V., Nishino K., Sato Y. (eds) Asian Conference on Computer Vision (ACCV). Lecture Notes in Computer Science, vol 10114, pp 342-359, 2016. (Published)

- Arigbabu O.A., Iman Y.L., Abdullah N., Helmee M. (2017). Computer vision methods for cranial sex estimation. IPSJ Transactions on Computer Vision and Applications, 9 (1),19 (Published)

- Arigbabu O.A., Iman Y.L., Abdullah N., Helmee M. Novel Group Variable Selection for Salient Skull Region Selection and Sex Determination. In: Ren J. et al. (eds) Advances in Brain Inspired Cognitive Systems. (BICS). Lecture Notes in Computer Science, vol 10989, pp 248-259, 2018 (Published)

- The Role of Group Variable Selection on Learning Important Skull Regions for Cranial Sex Estimation - Pattern Analysis and Applications (To Be Submitted)

Chapter 2

Literature Review

> "If I have seen further, it is by standing upon the shoulders of giants"
>
> Sir. Isaac Newton

This chapter provides a review of literature on the advances in sex estimation, both from the perspective of forensic anthropology and computer vision. The chapter begins with a detailed discussion of the conventional forensic scientists' approach to prediction of sex. Then, the classic and recent advances from the perspective of computer vision scientists are pointed out. Furthermore, the best performing techniques in each research domain in terms of the reported prediction results are highlighted. Finally, a discussion on the differences and shared similarities between the two domains is performed, and the possible ideology that could be established to integrate the two differing research domains is provided.

2.1 Cranial Sex Estimation in Forensic Anthropology

The use of pelvic region of human skeletal remains has often been studied by forensic experts to extract information about sex (Bidmos et al., 2010; Zech et al., 2012). This is due to their confidence in attaining reliable prediction, since the pelvic region shows clear morphological characteristics as a result of the functional demand of childbirth. However, there are a number of cases where the complete skeleton can not be retrieved due to the degree of damage caused to the subject or as a result of decomposition. The human skull has gained more attention in forensics and regarded as the second best indicator of sex (Spradley and Jantz, 2011). Moreover, in a high number of situations, the complete skull is usually available and it contains hard tissue whose sexually dimorphic characteristics can be preserved using suitable methods. Besides, it has been shown that the craniofacial region possesses some discriminant characteristics that can differentiate

between individuals (Ramsthaler et al., 2010, 2007). The mandible of the cranial is another region which has been independently studied by forensic anthropologists for sex estimation (Giles, 1964; Steyn and Ican, 1998). The examination of the mandible alone, however necessitate detaching the dental part from the cranial (Dong et al., 2015). Since this thesis does not focus specifically on the mandible due to the computational effort required to segment the region from the cranial, this chapter strictly concentrates on the literature reporting on the holistic use of the skull region for sex estimation. For the sake of simplicity, the case of mandible and cranial is literally analogous to iris and face biometric recognition, where there is a large body of research studying iris independently, yet the eye region is still predominantly present when a face image is analysed for face recognition purpose.

In forensics, both morphological and morphometric discriminant approaches are usually taken into consideration when assessing the cranial region (Franklin et al., 2012; Lo and Chen, 2003; Zech et al., 2012). In morphological assessment based methods, the analysts use their subjective perception (guided by their wide experience in the differences between male and female morphological variations) to identify sex from the skull, while measurement based or morphometric analysis method basically involves calculation of the geometric difference between a predefined set of landmark points.

2.1.1 Approaches based on Morphological Assessment

Morphological assessment is performed by visually making subjective inference on skull parts and determining their suitability in prediction of sex. A number of cranial traits such as orbital shape, frontal and parietal eminences, the slope of the forehead, and mandibular gonial eversion have been studied for sex determination. Earlier works in this aspect had relied solely on anecdotal comments by forensic anthropologists such as Hrdlička (1939) and (Comas, 1960), until the empirical study conducted by (Krogman (1962); Stewart (1979) ref. in (Konigsberg and Hens, 1998)). Krogman (1962) assessed 750 skeletal remains and claimed an estimation of 92% using skull alone. Stewart performed similar experiment with an estimate of 77%. However, it is worth noting that Krogman later stated that there was a bias in the sample distribution with a ratio of 15:1 for male. As such on an evenly distributed sample size the result could have reduced by 5-10%.

In the 90's, 5 main regions of the skull have been thoroughly studied and standardized on a 5-point ordinal scale to aid quantification of the sexually dimorphic traits that can be visually assessed (Buikstra and Ubelaker., 1994). These traits are the robusticity of the nuchal crest, size of the mastoid process, sharpness of the

supraorbital margin, prominence of the glabella, and the projection of the mental eminence. Generally to determine sex from a cranium, the traits are assessed and judged based on their similarity to the visual description provided in (Buikstra and Ubelaker., 1994), as shown in Fig.2.1.

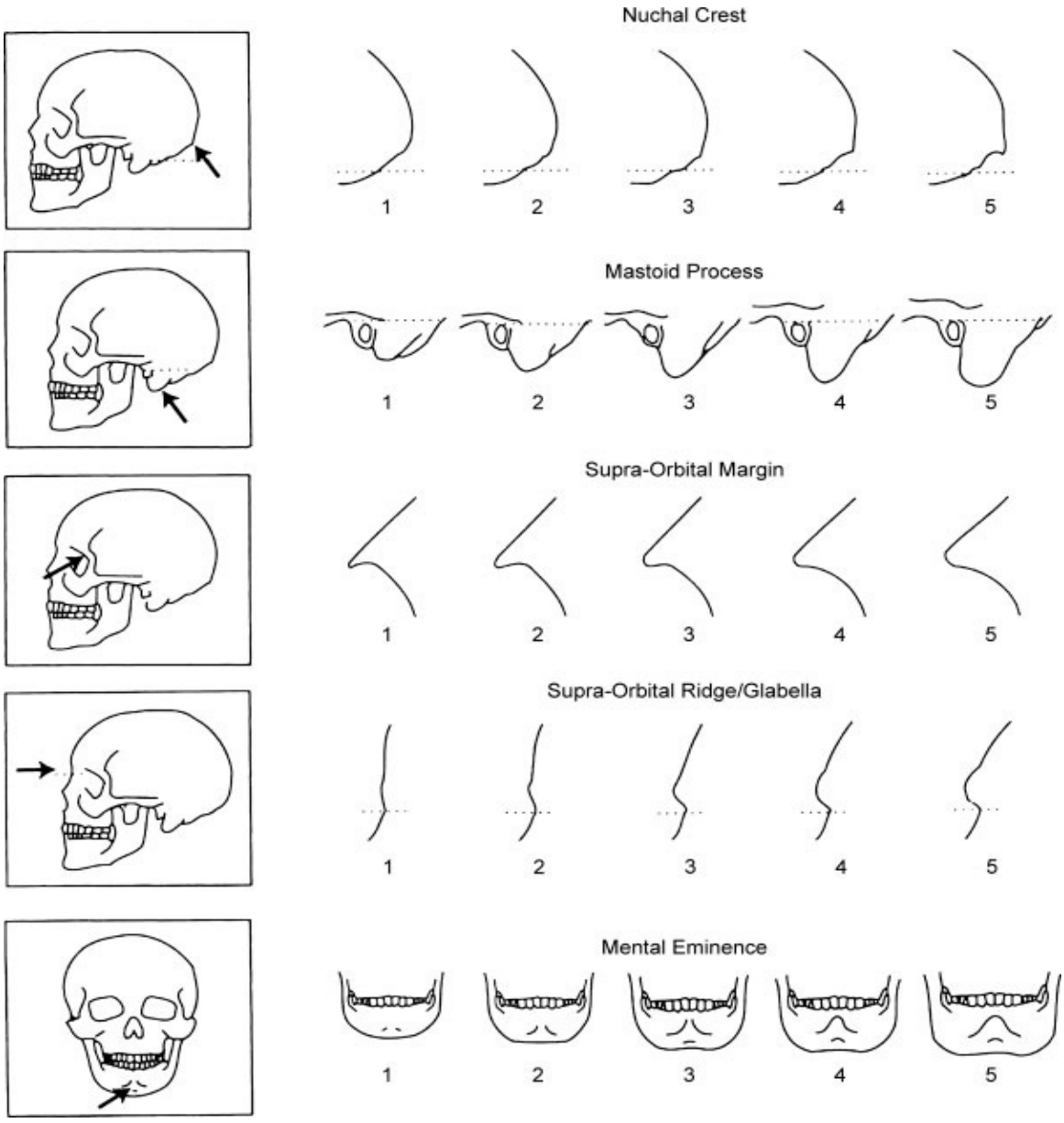


Figure 2.1: Standardized scale and diagram for scoring cranial traits (Buikstra and Ubelaker., 1994)

These ordinal scales from 1 to 5 can be interpreted as a range from "male" - "probable male" - "indeterminate" - "probable female" - "female". Although, the diagram has been used as a reference for assessing sexual dimorphism, there was no explanation by forensic anthropologists to justify the criteria for selecting these recommended traits and more importantly the visual descriptions used to explain

the traits are not comprehensive enough (Rogers, 1991). In his research, Rogers (1991) showed that variability exists in the number of traits, and visual descriptions used to explain those traits by studying 17 commonly used dimorphic traits. From the study, the author was unable to establish the criteria used by forensic anthropologists to make conclusion on the sex determinants due to unavailability of data on precision and accuracy of each trait.

Konigsberg and Hens (1998) visually assessed these 5 traits on a dataset of 138 samples using ordinal scoring of the traits. Motivated by the 5-points scale recommended in (Buikstra and Ubelaker., 1994), the authors collapsed the scale to 3-points to place the traits into 3 categories after being assessed by two different observers. For prediction, both logistic regression and probit models were adopted, and they attained a correct prediction rate of 83%. In another approach, the shape of supraorbital margin was assessed by Graw et al. (1999) on 108 skull data. The aim was to study whether only this feature could predict sex from the cranial. Plasticine impression was made to form the contour shape of the supraorbital and assess the shape of trait on a 7-point scale. The 7-points indicated the extent of expression of the supraoribtal curvature and its correlation to sex prediction. Their experiment presented a prediction rate of 70%. This indicates that the supraorbital margin is a suitable sexually dimorphic trait. Likewise using a single trait, Jung and Woo. (2016) examined the shape and size of mastoid process for sexual dimorphism using geometric morphometrics on 100 male and 100 female modern white American samples. The authors computed the centroid size and shape variables through thin plate splines (TPS). The leave one out cross validation was performed to classify the measurements into two classes. It was discovered that mastoid size may be a more significant characteristic for classifying each sex, with prediction of 87.3%.

However, when multiple traits are assessed, the prediction rate usually improves significantly. Williams and Rogers (2006) examined the accuracy and precision of 21 morphological traits on 50 adult skull data of White European ancestry between 32-94 years of age. By assessing the skull, the cranial data were score as male, female, or intermediate based on the presence or absence of the traits. An intra-observer assessment was performed to find the precision of the traits with respect to the sex identification results. The traits were ranked according to the lowest intra-observer error, and any trait with $> 10\%$ intra-observer error was discarded from further analysis. It was discovered that 6 traits (Mastoid, Supraorbital ridge, Size and architecture, Zygomatic extension, Nasal aperture, and Mandiblegonial angle) possess the highest precision and accuracy for sex prediction with a result

of 94%. Quite interestingly, the results of their experiment show that the 5 traits commonly used by forensic experts are among the 6 traits discovered in their study, which also solidifies the reason for choosing the aforementioned traits. The other trait, zygomatic extension indicated more bias in prediction accuracy, as it tends to often predict the skulls as male.

Walker (2008) presented detailed and informative analysis of these sexually dimorphic traits using 304 skull data, consisting of European Americans, African, and Americans with English origin. Predictive models such as k-nearest-neighbour, linear, logistic, and quadratic discriminant analysis were adopted for estimation. To make the traits meaningful for classification, the author modified the 5-point scale scoring system developed in (Acsádi and Nemeskéri, 1970; Broca, 1875) to cover an evenly distributed scale interval. The assessment was performed by two independent highly experienced observers, who rated the cranial region on their closeness to the examples presented in a diagram. There was negative bias in grading between the observers for all traits except the glabellar region. From the experiment, Walker found the combination of orbit, glabellar, nuchal, mastoid, and mental provided better indication of sex than any other traits combination, with a prediction rate of 90.1%. It was also shown that adding information on age and population produced little ($< 2\%$) increase in the prediction rate of the discriminant functions. His study indicates that sex is independent of age and local population. A similar study was carried out by Stevenson et al. (2009) on 304 skulls of English, African American and European American origin. But the discriminant function analysis was replaced with chi-square interaction detection (CHAID), as it was claimed that CHAID provides better explanation of the interaction between traits than discriminant function analysis. CHAID uses decision trees to maximize the probability of correct sex determinations. 31 different combination of the 5 dimorphic traits was performed, with 75% prediction rate.

In a study conducted by Garvin et al. (2014), it was shown that population indeed has a significant effect on the reliability of the dimorphic traits in sex determination (although age and body size showed no significant effect). The authors based their study on the standardized cranial dimorphic traits with the 5-points score scale of Walker. The examination was performed on 499 samples covering 4 different population groups (U.S. Whites, U.S. Blacks, medieval Nubians, and Arikara Native Americans). Their analysis produced population-specific estimates between 74% and 94%. It was noticed that the U.S. White samples exhibit more masculine and dimorphic glabella/supraorbital regions than the U.S. Black samples. Moreover, the glabellar and mastoid were found contributing more to sex

estimation than the nuchal. Similar observations were noted by Krüger et al. (2015) in the examination of South African samples. When Walker's logistic equation was applied the prediction rate reduced drastically to $68\% - 80\%$ for various combination of 3 to 5 traits. This is indicative that sex determination should be population-specific. A very successful method developed on a population group may not generalize to different population. Recently, Lewis and Garvin (2016) revisit the reliability of Walker's method subject to his outcome on intra- and inter-observer errors. Both errors are crucial in gaining intuition about the influence the population difference on sex estimation. According, Lewis and Garvin (2016), the intra-observer error rate of Walker may have been bias, as the author did not randomly select the samples used in the analysis, since 10 pre-selected samples were consistently used for calculating the score. Likewise, the inter-observer error rate may have been influence by the laxity given to the observers to ask the researcher designer some question before conducting the experiment. As a result, Lewis and Garvin (2016) re-calculated the approach of Walker on a sample of U.S white and black individuals using Cohens weighted kappas, intraclass correlation coefficients, and basing the examination on the ability of the observer to assign sex to the skulls based on their overall *gestalt*. The authors noticed slight disagreement in the scoring by the observers, which resulted in lower error rates in comparison to those reported by Walker.

Aside classical forensic approaches, modern computer-aided technology such as computed tomography (CT) was used to acquire the cranial data and visualize them via 3D reconstruction (Luo et al., 2013a,b). The authors proposed using sparse principal component analysis for sex determination from 188 CT scans of Chinese cranial samples. The CT scans were reconstructed into 3D mesh and the back part of the cranial was segmented out to reduce the number of vertices, focusing only on the frontal region of the skull. Then, the 3D coordinates from the craniofacial were concatenated and projected to the eigenface space. LDA was adopted for classification with accuracy rate of 91.3% and 90.1% for females and males respectively. In the same context, Pinto et al. (2016) in their recent work explored the sexually dimorphism of the supra-orbital region from 3D reconstruction using 2D wavelet for analysis. It was shown that such computer-aided technique could assist forensic expert in the morphological assessment of the skull.

2.1.2 Approaches based on Morphometric Analysis

This method has a long history in cranial osteobiography and tends to exhibit comparable sex prediction to visual morphological assessment. It relies on stan-

dard anatomical landmarks on the cranial data. The traditional approach of this technique to sex prediction was performed by computing different linear measurements (distances, angles, and/or ratios of distances) between the landmarks and evaluation using discriminant function analysis. The initial investigation on using craniometrics for sexing from skull was presented in (Giles, 1964; Giles and Elliot, 1963) where combination of 9 cranial traits measurement was used for sex determination with discriminant function analysis. Paiva and Segre (2003) performed sex prediction by evaluating the mastoid process using the calculated area of triangular distance between the Porion, Asterion, and Mastoidale. The areas were computed in two regions of the skull: left area, right area, and finally the total area of the left and right regions. The authors found an overlap of 60% in the right area of male and female, and 51.67% overlap in the left area. Using the total area, the overlap was 36.67%. A similar cranial assessment method was proposed in (Galdames et al., 2008) on 81 Brazilian cranial data. The mastoid region is very resistant to physical damage and usually remains unaffected, even in a very old or damaged skull, because of the anatomical part of the cranial it is positioned and its compact structure (Kanchan et al., 2013).

While there are methods using the traditional linear morphometrics, it has been claimed that linear measurements do not totally capture the difference in the shape of some traits such as the prominence of chin or glabella (Gonzalez et al., 2011). Besides, the method can easily be influence by changes in size of the cranial and it is population specific. Even in the same population, the measurements collected in the first half of the 20th century cannot guarantee reliable prediction on recently collected cranial samples, particularly with the rate at which globalization is constantly increasing (Bigoni et al., 2010). To this end, forensic anthropolgists have studied the use of geometric morphometrics (GM), particularly from digitized 3D CT data, which in addition to the linear measurements provides the 3D geometric structure and coordinates of anatomical landmarks (Adams et al., 2004; Bookstein, 1997).

The standard landmarks that can be used for sex determination are shown in Fig 2.2. The traits are: Ectoconchion (ec), Superior orbital margin (os), Inferior orbital margin (io), Fronto-zygomatic orbitale (fo), Fronto-malar anterior (fma), Alare (al), Zygomaxillare anterior (zm), Dacyron (d), Nasomaxillary (nm), Ectomolare (ecm), Jugalia (ju), Auriculare (au), Porion (po), Asterion (ast), Foramen magnum lateral (fml), Mastoidale (ms), Zygion (zy), Nasion (n), Glabella (g), Nasospinale (ns), Prosthion (pr), Lambda (l), Bregma (b), Opisthion (o), Basion (ba), Frontal convexity (fc), Parietal convexity (pc), Occipital convexity (oc),

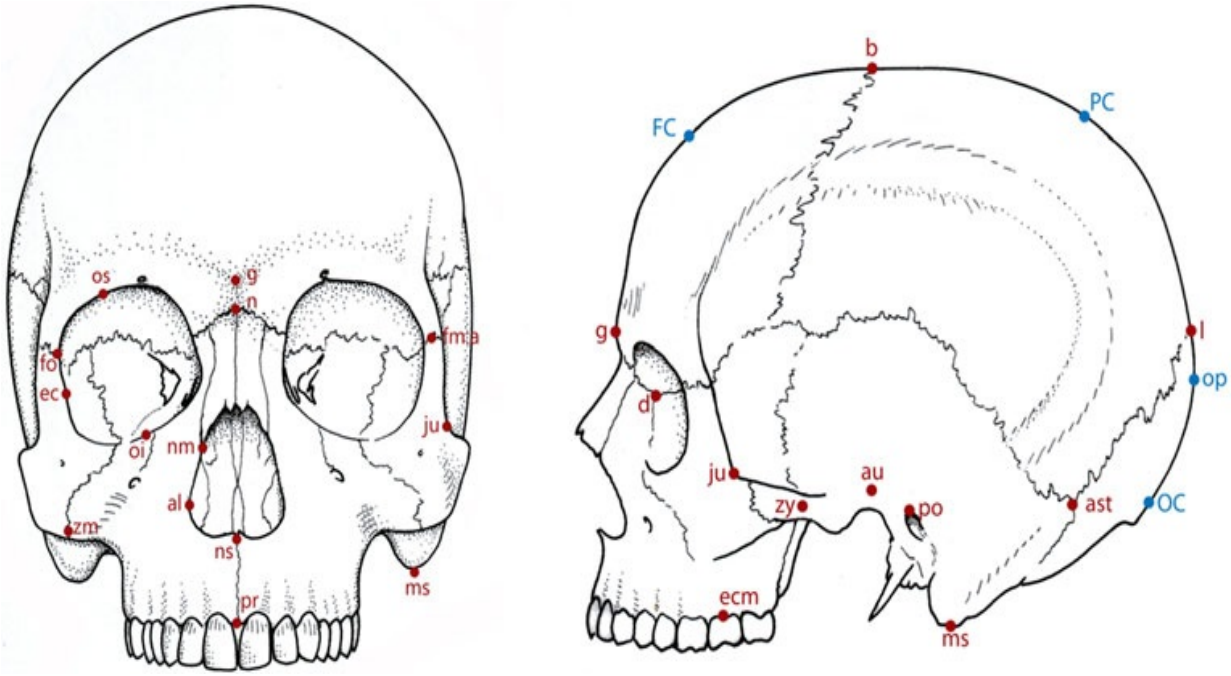


Figure 2.2: Standard landmarks of sexually dimorphic traits of the cranium (Franklin et al., 2012)

.

Opisthocranion (op).

Most of the morphometric methods analyse shape with procedures based on Kendall's shape space, Procrustes distance, or their tangent space approximations (Adams et al., 2004). Forensic analysts can preserve the relative location of each landmark point and evaluate the multivariate result of variation in sexual dimorphism as configurations of landmarks back in the original space of the cranium, and assess variation in shapes with few or no landmarks (Adams et al., 2004; Gonzalez et al., 2011). A number of in-depth studies have been conducted in this aspect of craniometrics and some solutions have been formulated to indicate the discriminant cranial regions that contribute more to sex dimorphism. Moreover, the authors have highlighted the influence of size on the shape of the cranial of males and females.

Franklin et al. (2005) performed morphometric analysis of 8 dimensional measurements of 3D landmarks from 332 South African samples. By using discriminant function analysis, it was found the facial width (bizygomatic breadth) and the length and height of the cranial vault possess higher sexual dimorphism. The accuracy rate of the system was between $77\% - 80\%$. Furthermore, it was confirmed that male cranial is generally larger than female cranial. Other studies have found additional regions expressing greater sexually dimorphism. Bigoni et al. (2010)

presented a craniometric approach to sex determination with the aim of locating the regions of the cranium where sexual dimorphism are most pronounced. The study was conducted on 139 cranial set of Central European population. 82 ecto-cranial landmarks were annotated from 7 regions (the configuration of neurocranium, cranial base, midsagittal curve of vault, upper face, orbital region, nasal region, and palatal region). The generalized procrustes analysis (GPA) was adopted to performed shape analysis and thin plate spline was used for visualizing the deformation of shape to understand which landmarks were responsible for shape difference. They noticed no sex difference in the sample set when landmarks from the whole cranium were used. As a result, further partial shape examinations were conducted on 7 sub-regions of the cranium. It was discovered that the cranium base and the neurocranium configuration did not depict any sex variation, but the midsagittal curve, the upper face, the orbital region, the nasal region, and the palatal region possess strong sexual dimorphism.

Franklin et al. (2012) examined the morphometrics of 200 Australian adult (100 male and 100 female) CT scans. The authors computed inter-landmark distance between 46 anatomical landmarks such as Nasospinale, Mastoidale, Glabella, Frontal Convexity and so on. Also included was the centroid size (the square root of the sum of squared distances between the landmarks and their centroid) and shape variation of the landmarks using procrustes analysis. Using only shape and centroid size they attained accuracy rate of $83.5\% - 88\%$. When incorporated with linear measurements the accuracy rate was between 87.5% and 88%. Jiménez-Arenas and Esquivel (2013) presented a performance comparison of two discriminant function analysis by measuring 17 craniometric variables. It was discovered that most metric variables were higher for males than females. Rosas and Bastir (2002) have studied the influence of size and shape of the cranium on sex determination on a sample set of 52 males and 52 females within the ages of 21-50 years. 29 standard 3D landmark coordinates were collected from the cranial data. The thin-plate-splines (TPS) was used for warping the points onto the target skulls for morphometric analysis and multivariate analysis was performed using MANCOVA generalized linear model. They discovered that there was no influence of size on shape in the two sexes.

Though several researchers have suggested that size does not influence the shape of the cranial, it should be noted that these techniques were mostly applied on specific populations. An example of a situation where population-specific methods fail to generalize to different population is the evaluation performed on the standardized demographic prediction software, FORDISC. Normally, any combination of

standard measurements can be used for sex identification using multi-discriminant analysis in the FORDISC system (Ousley and Jantz 1996). The performance of FORDISC 3.0 originally modeled with data samples from 28 different populations, was recently tested on samples collected from France and Thailand. It was concluded that while the application would normally predict sex with an accuracy rate between 74% and 86%, it could not generalize to these two independent populations. For example the sex prediction rate dropped to 52.2% and 77.8% respectively for the two populations (Guyomarc'h and Bruzek, 2011). This is part of the limitation of morphometrics, as the discriminant functions are usually population specific. Besides, a recent comparison of predictive models such as linear discriminant analysis, linear regression, and support vector machine on craniometric landmarks showed that the methods developed on Western samples could not predict accurately on Thai samples (Santos et al., 2014). In fact, as mentioned by Franklin et al. (2013) it is preferable to build population specific standards with proper documentation of skeletal collections, which can serve as repository for acquiring the required osteometric data, when performing sex determination.

This is the context provided by Mehta et al. (2014) where craniofacial parameters of 143 samples from Gujarati population composing of 66 female and 77 males were analysed. The authors considered measurement between anatomical landmarks such as maximum frontal breadth, minimum frontal breadth, maximum cranial length, morphological facial height, nasionprosthion length, bizygo- matic breadth and bimaxillary breadth and discriminant function analysis was used for determining the sex of each skull. They obtained prediction results between 61.3 % - 88.7 % in the study, with the maximum cranial length and bizygomatic breadth providing the best individual accuracy of 78.2%. Similarly, 70 adult samples consisting of 43 male and 27 female of Indian population were analysed by Ramamoorthy et al. (2015) using 26 craniometric parameters with discriminant function analysis. The analyses were done using stepwise, multivariate, and univariate methods with accuracy of 77.1%, 85.7%, and 72.9% respectively. In addition, a study of the outline of the shape of the cranial in form of geometric morphometric analysis has been conducted to understand its effect on population group (U.S white and black) and sex (Murphy and Garvin, 2017). In order to perform the analysis, elliptical fourier analysis (EFA) was utilized to extract features from the 2D projection of the 3D cranial, mainly capturing the left lateral, posterior, and superior cranial views, eventually resulting in 80 EFA coefficients. Then, PCA was performed on the coefficients and two-way MANCOVA was computed on the set of male and female outlines to assess the sex and population group differences. Also, ANOVA

was computed on the significant principal components to understand which parts of the shape outline contribute to the group differences and finally DFA was used to predict the class of each sample. The experiment resulted in a sex prediction rate of 70.2%, indicating that outline alone may not be a comprehensive form of cranial representation.

Summary

Due to the ever increasing improvement on computer technology, the morphometric based method have often been adopted instead of morphological methods through the use of the discriminant function analysis. Despite this advancement, forensic anthropologists stills rely heavily on visual assessments of sexually dimorphic traits. Some of the reasons for morphometrics not being able to totally replace subjective visual assessment include the large amount of time and expenses required to compile anatomical landmarks from cranial data. In addition, forensic experts and anthropologists usually face difficulty to accurately perform measurements that are sufficient enough to cover these sexually dimorphic characteristics, which can easily be visually perceived (Graw et al., 1999). Also, morphometric analysis can be distorted by population variation in cranial size.

Based on this variation effect, visual assessment can still provide better understanding/interpretation of the difference in expression of sexually dimorphic traits. Nevertheless, visual assessment to a larger extent is insufficient to independently provide accurate information about sex, because there are evidences showing the implication of inter-observer subjectivity on cranial visual assessment. This discrepancies eventually have a high statistical significance and the results of such analysis as mentioned by Walrath et al. (2004) should be population-specific and interpreted with caution (Lewis and Garvin, 2016). Therefore, it can be concluded that using the conventional forensic anthropologists approach, sex determination in population specific distribution provides more confident estimates than in non-population specific. Moreover, both morphological assessment and morphometric possess some limitations such as subjectivity, requirement of high amount of time for landmark localization, and influence of skull variation on generalization on metric traits.

2.2 Gender Recognition in Computer Vision

There is a growing interest in the domain of computer vision in automatic classification of gender from human face images. This is due to the emergence of several electronic applications such as human-computer interaction, surveillance, access

control and market analysis (Sun et al., 2018). In this review, previous works on gender recognition are reported based on the feature extraction techniques and estimation models adopted.

The problem of face gender classification involves 4 main stages: face detection, preprocessing, feature extraction, and classification (Gonzlez-Briones et al., 2018). Face detection is the stage where the face region is localized and segmented from the unnecessary background region. Several algorithms have been proposed for face detection which have very low false positive rate on 2D face images (Viola and Jones, 2004; Wang and Yuan, 2001) and video (Hjelmås and Low, 2001). After face detection, the face image can be preprocessed to remove noise, normalize the illumination, and align to centralize local regions like eyes, nose, and mouth. Feature extraction is the stage where discriminating feature are extracted to differentiate male from female and classification is the stage where predictive models such as nearest neighbour, linear discriminant analysis (LDA), support vector machine (SVM), artificial neural networks (ANN), and so on are used for determining the gender group of the face image.

From the literature, two mainly approaches have been considered for extracting face features and classifying gender from 2D and 3D images, which are: geometric based and appearance based methods (Abbas et al., 2018; Arigbabu et al., 2015; Khan et al., 2011).

2.2.1 Geometric based

This technique finds some salient points on the face and measures the geometric distance between those points to form a feature vector.

Fiducial Distance: is an example of geometric based feature extraction. In this method the local regions of the face such as eyes, mouth, nose, chin are marked with a number of landmark points describing the features around those regions. Then, the fiducial distance can be obtained by performing several geometric measurements between the marked points. This method has been shown to be effective in gender recognition. Brunelli and Poggio (1992) computed 18 point-to-point distance to train a hyper basis neural network, they attained a recognition rate of 79%. Burton et al. (1993) utilized the distances between points on the face as feature vectors and linear discriminant analysis for gender classification. Fellous (1997) measured 22 normalized vertical and horizontal fiducial distances from 40 facial points.

With the assistance of human operator, they were able to manually annotate these points from images of frontal faces as shown in Fig 2.3. From the normal-

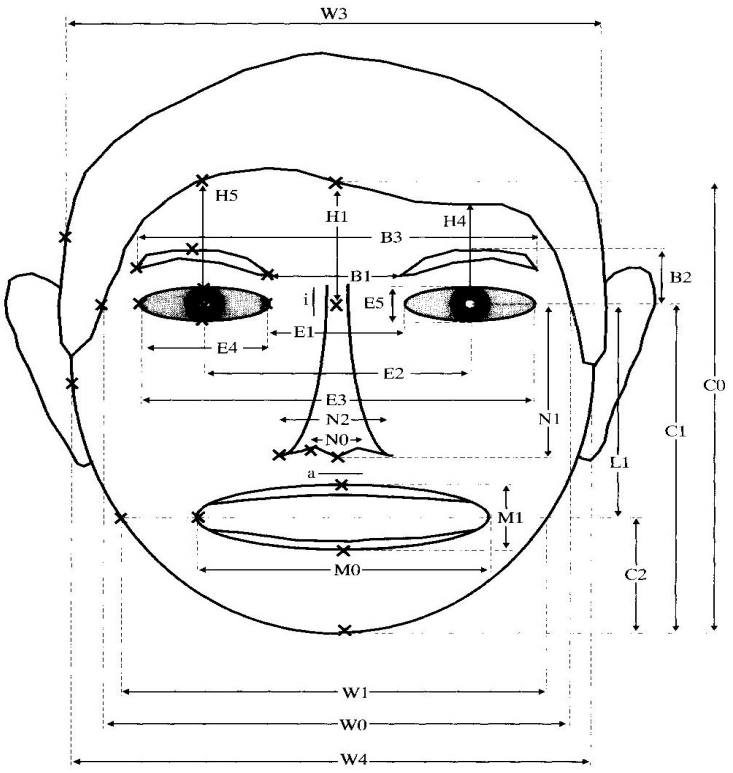


Figure 2.3: Face measurements performed by Fellous (1997)

ized distances, five dimensions were derived using discriminant analysis to classify gender with a recognition rate of 90%. Such geometric feature extraction concept has also been extended to 3D mesh representation where prominent facial features which discriminate between the genders can be extracted by manually placing landmarks on specific areas of the face. Han et al. (2009) used the approach to define shape features of the eyes, nose, mouth and cheek, as shown in Fig 2.4.

The features were measured in terms of the ratio of that region on the surface to the whole face. By using SVM with RBF kernel for classification, an overall error rate of 17.44% was attained. Hu et al. (2010) extracted the landmark points centring broadly on 4 specific facial regions. The landmarks were detected automatically by using the profile curvature projected along different axial view (x- and y-coordinates). On detecting the landmarks the face was partitioned into 4 regions (upper face, lower face, left eye, and nose region). SVM was used for classifying each region and the results of 4 regions were fused using a weighted sum fusion method. An accuracy of 94.3% was attained which was better than the result of each independent region. Furthermore, the use of radial curves to select and highlight the most salient face region that contribute to gender recognition has been suggested in (Ballihi et al., 2012). Based Riemannian shape analysis, the Euclidean

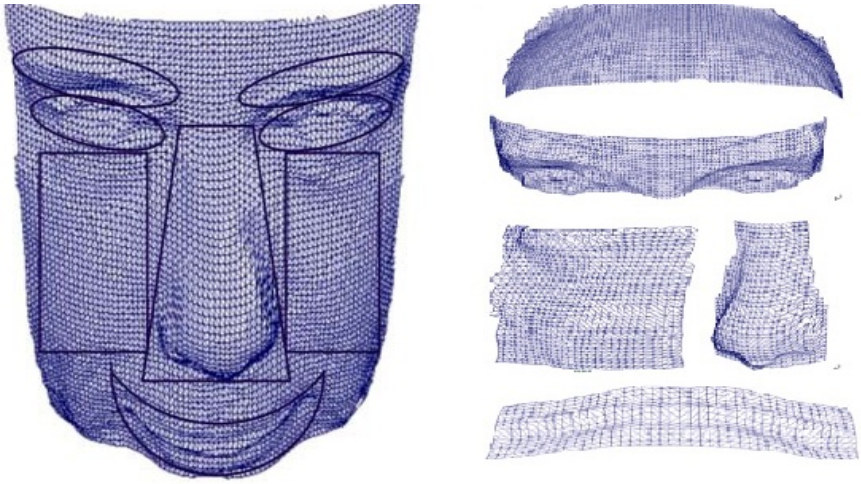


Figure 2.4: Extraction of primary features of the face (Han et al., 2009)

distance function from a starting point (mostly placed on the nose point) to different region of the face surface can be computed to collect a large set of geometries curve features. By using Adaboost as a feature selection scheme a recognition rate of 86% was attained. Using statistical methods, Yang et al. (2017) introduced 3DPCA and 3D independent component analysis (3DICA) for gender recognition on 3D face landmarks and SVM for classification. Recently, a new curvature based feature extraction method which computes the geodesic path between landmark points was introduced by Abbas et al. (2018). Since shape is an important characteristic description of the face, particularly for distinguishing between 2 genders, the authors suggested that capturing the 3D geodesic path curvatures could be useful information for gender analysis. After computing the features, LDA was adopted for classification and a result of 87.3% was attained. Combination of the new representation with conventional euclidean distance inter-landmark measure produced an improved result of 88.6%.

2.2.2 Appearance based

This method makes use of the holistic intensity, as well as, statistical distribution of the face. It does not necessitate localization of specific landmark points on the face, since the entire information of the face can be quantified by assessing the probability distribution of pixel intensity. Some of the commonly used appearance-based techniques are discussed as follows:

Pixel Intensity Values: the raw pixel intensity values of an image can be directly used to train a classifier such as SVM or ANN. The method was used in one of the earliest approach on gender recognition Golomb et al. (1990). The author proposed

SEXNET based on back-propagation neural network classification (BPNN) with 30 by 30 resolution face image. The technique attained a classification rate of 93% on only raw pixels. Tamura et al. (1996) used 8 by 8 and 8 by 6 low resolution face images for gender identification with 90 % accuracy.

Yang and Moghaddam (2000) used linear SVM for gender recognition with raw pixel and an error 6.7% was attained. Since, the entire dimension of the image will be input to the classifier after face detection and preprocessing, the length of the feature vector can increase as the size of image increases. Therefore, some authors have used dimensionality reduction techniques such as principal component analysis (PCA)(Bekios-Calfa et al., 2014; Dhamecha and Sankaran, 2011), independent component analysis (ICA), canonical correlation analysis (Shan et al., 2008) and their variants for to decrease the size of the feature vector prior to classification. These feature reduction techniques are based on statistical procedure . For example PCA exploits the joint probability distribution of the correlated data to find an orthogonal transformation that can maximize the variance between those data points. These orthogonal transformation properties are the eigenvalues and eigenvectors used to project the data point to new uncorrelated direction/space.
Wavelet Transforms: are feature extraction methods that transform the spatial distribution of an intensity image into frequency domain where coefficient from wavelet sub-bands can be used to representation the important information of the face. Wavelet transforms such as discrete cosine transform (DCT), discrete wavelet transform (DWT), fourier transform have been explored for face gender recognition (Berbar, 2013; Khan et al., 2011; Majid et al., 2003).
Local Feature Descriptors: in recent years, the challenge of face images have grown bigger with the availability of unconstrained images from web, video surveillance, and mobile devices. As such, use of local feature descriptors have attracted more interest for gender recognition because the features can provide comprehensive local and global representation of the face under various imagery challenges such as illumination, pose, expression, and occlusion. Examples of local feature descriptors are local binary pattern (LBP) (Ojala et al., 2002), histogram of oriented gradient (HOG) (Dalal, 2006), scale invariant feature transform (SIFT)(Lowe, 2004), and weber law descriptor (WLD) (Chen et al., 2009).

SIFT is a keypoint descriptor that uses the difference of gaussians (DoG) at various scales to detect the dominant pixel locations (keypoints) in an image. Then, it extracts histogram of gradient orientations by considering the neighbouring pixels around the keypoints. It has been used for face gender recognition in 2D and video sequence (Demirkus and Toews, 2010; Wang et al., 2010). LBP converts the image

into strings of binary numbers to encode texture descriptions of the face. It is invariant to monotonic gray scale changes and has been found very effective for face gender recognition, especially when combined with adaptive boosting (Adaboost) to select the most important features (Shan, 2012). WLD is based on weber's law, and the descriptor performs non-linear quantization of local intensity information from multi-scales. Ullah et al. (2012) applied block based WLD for gender recognition with distance metric based classifiers such chi-square and euclidean distance and an accuracy of 99.08% was reported on the well known FERET face database. HOG encodes shape representations by casting votes of gradient magnitude from local cells into histogram bins based on the local orientation of the image gradients. HOG has been widely applied for face gender recognition (Alexandre, 2010; Modesto et al., 2013) and its variant pyramid histogram oriented gradient (PHOG) showing similar performances (Arigbabu et al., 2014; Khan et al., 2014).

There are other methods such as the sparse representation technique proposed by Yang et al. (2013a) where the standard dictionary learning which is based on using class-specific fixed atoms was improved to learn discriminatory dictionary that is able to adaptively learn latent information among class labels. Also, the deep learning approach proposed by Mansanet et al. (2015), which is able to learn local information from the input image and use discriminative deep network to classify the learned features. This has also created the avenue for researchers to explore the possibility of extending face gender recognition to cater for large scale dataset. For instance, Jia et al. (2016) trained deep convolution neural network (CNN) on five million weakly labeled face images, with several variation on the architecture (depth) of the network and the results were compared with SVM based methods. The results of CNN was better than SVM with accuracy of 98.9% and able to generalize across different dataset i.e a separate dataset was used in training, while different one for testing with accuracy of 91.34%. Rodrguez et al. (2017) studied the face analysis problem "in the wild", showing the possibility of age and gender identification under highly challenges conditions such as occlusion, expression, illumination, and deformation. Normally deep networks exhibit sensitive to such variations, however the authors proposed an attention mechanism which is able to automatically discover the most informative and reliable parts of a input face, limiting the effect of variation on the prediction result. The approach showed promising results with 2.3% improvement over the state-of-the-art reported in the literature.

From 3D perspective, some studies analysing the importance of face appearance have also been reported. For instance, Xia et al. (2013) investigated the

relation between face symmetry and gender by capturing the facial asymmetry using Deformation Scalar Field (DSF). These extracted information from each 3D face are transformed to a lower dimensional space using PCA, which are then fed to predictive models such as Random Forest, Adaboost and SVM. They attained a classification rate of 90.99% with Random Forest, and their examination showed that there is a positive relationship between gender and facial symmetry. This approach has been extended to examine the relationship between face averageness and symmetry of 3D face image for gender recognition (Xia et al., 2015). However, instead of transforming the DSF features with PCA, correlation based feature selection was adopted and the recognition results improved to 93.78%. Huynh et al. (2013) proposed computing local binary pattern (LBP) on RGB-D depth image and SVM for classification. Their experiment produced better classification results than standard LBP and 3DLBP.

2.3 Relation between Forensic and Computer Vision Methods

In the previous sections, summary of recent and classic works on sex and gender estimation both from the perspective of forensic anthropology and computer vision respectively have been presented. Although, the two field are miles apart, it is worth pointing out that they both loosely share similar research orientation, yet have differing conceptual ideology. In forensics, two main approaches have been explored to date for sex estimation: morphological assessment (non-metric) and morphometrics (metric) based techniques. From the literature, it has been pointed out that morphological or visual assessment suffers from subjectivity and at times it is highly difficult to find inter-observer agreement. An example is in the case where the experience of an observer is very important, as they estimate with high accuracy in a population group they are familiar with but tend to falter in an unknown population dataset. Similarly, morphometrics which is based on linear or geometric measurement quantified with mathematical formulas tends to be affected by the changing size of the cranial. Moreover, it requires a lot of human effort and precision to make large collection of accurate anatomical landmarks. In computer vision, researchers have used both visual assessment from human observer and feature representations to make assessment of gender from face image. The visual assessment aspect can be slightly considered to have connections with forensics, since the main aim is to gather observation from an observer with regard to the observer's perception. In the feature representation method, two techniques are commonly used: geometric representation and appearance representation. The geometric representation makes use of fiducial landmark points both from 2D and

3D face data. Again, this can be considered to be slightly related to morphometrics in forensic examinations of the cranial, because the measurements have been establish based on the study of anthropometric development of the craniofacial (Farkas, 1994). Then again, geometric measures have similar limitations found in standard forensic anthropology research such as effect of change in face size. With regard to this, it can be literally suggested that these two research domains share similar research direction, albeit forensics being more rigorously and strictly investigated with well grounded principles. However, an aspect that has not been explored in forensic anthropology is representation of skulls with local feature descriptors, which from computer vision research perspective has exhibited tremendous success. Hence, part of the contribution of this research is to fill this research gap by exploring the possibility of using 3D shape descriptors for demographic estimation.

2.4 3D Local Descriptors and Patch based Representation

The essence of this section is to point out some of the advancement on 3D feature representations and their usage in computer vision applications. 3D representation is a standalone field with immense contributions from diverse perspectives. Hence, it is practically impossible to cover all the existing methods on 3D feature representation. However, since 3D descriptors are explored in this research, it is therefore reasonable to report the most prominent innovations on 3D feature representation methods. To this end, this review groups the body of research into 3 broad categories, which are: (1) Feature based (2) View based (3) Graph based descriptors, and describes the most commonly used techniques in each category.

2.4.1 3D Shape Representation for Object Recognition

Feature based descriptors

Feature based descriptors describe an object by encoding the distribution of local properties such as curvature, spatial maps, orientation, and pose in form of histograms and shape distribution. The descriptions can be extracted to target the global or local outline of the 3D objects, and there are cases where both local and global are embedded in the final representation. The classic methods to the 3D object retrieval focused on describing the global shape properties of the objects such as area, moments, and volume. Osada et al. (2002) used the global shape distribution to represent 3D models. The shape distribution was formed by transforming the 3D model into a parameterized functions, which can easily be compared with other models. Such functions include the angular distance between three random

points (A3), the surface distance between two random points (D2), the area of the triangle between three random points (D3), and the volume between four random points on the surface (D4). Ohbuchi et al. (2005) proposed an improvement on the D2 distribution, which captures information about surface orientation and surface distance. The method converts the 3D model into orientated points, then calculates a joint 2D histogram of the distance between two points and their mutual orientation as features. Other primitive functions of the 3D surface have commonly been explored for encoding shape distribution. For instance, Zaharia and Preteux (2001) have utilized the shape index which mainly reflects the angular coordinates of the principal curvature of a surface to extract the intrinsic shape description of 3D mesh. Global descriptors are not too computationally demanding because they can summarize the holistic description of the 3D object in a smaller dimension. However, they lack the capability to cope with small variation in the shape of objects from the same class, which eventually affects the retrieval performance.

Table 2.1: 3D Shape Descriptors

Category	Algorithm
Feature based	SIFT Spin Image LBP Shape Index Map Heat Kernel Signature (HKS) Scale Invariant HKS Spherical Harmonics (SH) PCA-SH Fourier Transform
View based	Light Field Descriptor (LFD) Compact Multiview Adaptive View
Graph based	Skeletal Graph Reeb Graph
Learning based	Sparse Representation Deep Learning

Alternatively, the global features can be derived using representation methods based on transform coefficients. Transform coefficients based descriptors can be computed by decomposing a 3D model according to a set of basis functions. These basis functions are formed to pack the main patterns for low order projection. This kind of representation is mathematically sound, because of orthogonality, completeness, and consistency (Cheng and Chung, 2016). Examples of transform

based descriptors are spherical harmonics and fourier transform. Spherical harmonics descriptor (SHD) is based on the fundamental theory of spherical harmonics where any spherical function can be decomposed into the sum of its harmonics (Kazhdan and Funkhouser, 2002). The descriptors is encoded by computing the 3D Fourier transform of a shape function defined over a set of concentric spheres in spherical coordinates coordinates on the voxel grid of the 3D object (Cheng and Chung, 2016; Kazhdan et al., 2003). One of the main advantages of spherical harmonics is that it is rotation invariant. In other words, the L_2-norm does not change despite rotating the spherical function. To further improve its rotation invariance ability PCA-SH has been proposed which uses PCA to normalized the 3D object before computing SHD (Vranic and Saupe, 2002). It was demonstrated that incorporating this step into the matching problem can improve the retrieval performance (Vranic, 2003; Vranic and Saupe, 2002). Another global 3D descriptor is heat kernel signature (HKS) which is based on heat propagation on a riemanian manifold (Sun et al., 2009). This descriptor and its variants have also exhibited tremendously efficiency in 3D object retrieval (Bronstein et al., 2010; Sun et al., 2009).

More popular trend in feature-based representation concentrates on describing a 3D model with local representation, because better discriminating characteristics can be derived according to the local neighbouring connectivity of the 3D object. Johnson (1997) introduced spin image which uses random subset of points from the 3D mesh to encode the angular difference between a centre point and its neighbouring points. Due to the immense success of SIFT descriptor on 2D image recognition, several authors have proposed various techniques to extend the approach to 3D object representation (Darom and Keller, 2012; Smeets et al., 2013). Likewise, the LBP descriptor which became the state-of-the-art descriptor for 2D face recognition, texture analysis, object retrieval in the mid-2000 has recently been extended to 3D mesh and its performance evaluated on a number of 3D object retrieval tasks and 3D face recognition (Werghi et al., 2015b). Unlike 2D where a binary ring around the centre point can typically provide information about the texture of the 2D object, in 3D LBP extracts binary representation of the curvature property of the 3D model. Thus, it reflects angular transition as well as the shape variation in a region of interest bounded by the size of such region.

View based descriptors

The descriptors under this category describe a 3D shape by how it looks in 2D projections. In other words, the descriptors thrive on the multi projection of the

3D shape from difference angles onto different 2D planes, before computing the representation. Methods of projection to 2D exploit properties such as silhouette, depth-buffer images, and grayscales. Some view based descriptors are: Light Field Descriptor (LFD), initially introduced by Chen et al. (2003) involves projecting the 3D objects to 20 different 2D views from different camera position. The camera positions are selected at the vertices of a regular dodecahedron. The result of each view is silhouette (binary) image. 10 of the 20 views are then taken from opposite vertices of the dodecahedron, since the image from the remaining views will be identical. For each selected image, a combination of 35 Zernike moments and 10 Fourier descriptors are used to represent the image. Finally, the similarity between two feature vectors is calculated using L1 distance. Compact Multi-View Descriptor (CMVD) is based on projecting the 3D mesh on 18 2-dimensional views, selected from the vertices of 32-hedron. The 2D image are represented with features obtained from polar-fourier transform, zernike moments, and krawtchouk moment. The resulting features are matched to compute the comparison between two 3D models, with the minimum sum of distance denoting the similarity between the models. View based methods are typically dependent on the assumption that two 3D models are similar if they look the same from different 2D views. In order to select the optimal number of views that are sufficient enough accurate matching, Ansary et al. (2005) introduced an adaptive view clustering (AVC) algorithm which uses a probabilistic bayesian method for 3D model indexing. Basically, as a 3D model becomes more structurally complex, the 2D projection of such model will have less similarity, thus the AVC technique enables selection of dominant 2D projections according to the geometric complexity of the 3D model. The methods uses an alternating k-means to cluster the 2D views and then include additional characteristic views where necessary before using the bayesian information criterion (BIC) to select the most representative views for the 3D model.

Graph based descriptor

This category of descriptor is different in computation compared to feature and view based descriptors. It uses the topological information of the 3D object to describe its shape and the graph that is constructed exhibits how the different components in 3D shape are connected together. Example of graph based descriptors are model, reeb and skeleton graphs. Despite these properties graph based descriptors usually lack the ability to generalize to different representation form and require specific matching technique. Moreover they are computationally expensive and easily affected by small topological changes (Zhang et al., 2004).

Learning based descriptor

Recently, computer vision scientists have suggested data driven method of representation. In a condition where there is a large enough dataset covering diverse data distribution, learning the representation directly from data has shown to be an effective method of deriving object representation. Deep learning has become a trend among CV researchers and CNN which performs a series of convolutional computations has been practically tested in object recognition applications (Wu et al., 2016). Unlike 2D CNN, some authors have demonstrated that 3D CNN incorporates spatiotemporal information in the learned features. The 3D CNN uses a $3 \times 3 \times 3$ convolution kernels in every layer, giving temporal depth to the extracted features. This type of architecture with a simple linear classifier at the top has produced performances which are better than state-of-the-art techniques on different benchmarks datasets (Du et al., 2015; Zhou and Tuzel, 2017). A different method of learning features from 3D data using deep networks is to consider working directly on the 3D point clouds, where the input to the network is set of unordered points. For instance, Qi et al. (2016) proposed the PointNet which accepts 3D point cloud into a recurrent neural network (RNN), which trains the data sequentially by randomly permuting the input point clouds to make network invariant to the order of the data. Then, it uses a max pooling layer as a symmetric function for feature aggregation. This sort of approach has further been modified to incorporate metric learning in the data sampling stage. The main idea is to divide the input data into overlapping local region by distance metric, then extract features which capture the local geometrical structure of the small local neighborhood. The extracted feature a later accumulated into large batch and processed to obtain high level features (Qi et al., 2017).

2.4.2 3D Shape Representation for Craniofacial Analysis

3D shape descriptors have been explored for analysing human craniofacial, although the application is mostly centre around the soft tissues. The studies in this domain typically focus on anatomical landmark detection and medical condition examination such as Deformational Plagiocephaly and 22q11.2 Deletion Syndrome. Hammond (2007) initially examined facial deformation by manually placing a number of landmarks on 3D mesh data to align the faces to an average face, similar to the standard dense surface method. Then, PCA was used to represent the entire dataset, and the coefficients were used to differentiate the facial morphology of children due to genetic disorder. Motivated by the work of Hammond (2007), the global outline of the frontal and side view profile of the head were extracted from

2.5D depth map for head shape classification in (Shapiro et al., 2009). Atmosukarto et al. (2010) extracted low level features based on standard shape primitives such as gaussian curvature, besl-jain curvature, azimuth-elevation angle to describe different types of head shape dysmorphology in children. Further, anatomical keypoints were manually selected and learned with SVM to determine additional salient keypoints that improve the head shape retrieval performance. Yang et al. (2013b) extracted the global outline of the head and sparse logistic regression classifying the severity of skull deformation. These automated techniques are useful digital diagnostics for clinicians to retrieval of CT images of patients with skull deformity.

2.5 Conclusion

Forensic methods which require accurate landmarks localisation, only select sparse landmarks to capture global shape, but fail to capture the local shape description of the skull. Meanwhile, gender classification based on 2D/3D face images rely on local textures or identification of key landmarks such as nose tip for constructing geometric descriptors. However, these information either do not exist or is severely lacking on skull data. Methods presented in 3D object classification are either part based or shape based, whereby 3D descriptors are summarised over local region. They are effective for classifying 3D objects that are significantly different in shape. However, it is not evident whether they would perform similarly for sex identification with skulls that do not differ significantly in overall shapes. On the other hand, it is clear from the literature that for sex/gender identification (using face or skull), the overall shape and morphological differences between the subjects of two opposite sexes play important roles for accurate classification. Thus the main aim of the thesis is to explore morphological representations of 3D skulls that take advantage of 3D object classification (e.g., 3D descriptors), and in the meantime overcome the issues of lacking local shape description (in traditional forensic methods) and issues of lacking meaningful texture on skulls and/or problems to easily identify the landmark corresponding to nose tip (which methods for gender classification from 2D/3D faces generally rely on).

Chapter 3

Research Methodology

> There are some things one can only achieve by a deliberate leap in the opposite direction.
>
> Franz Kafka

3.1 Overview

The base structure of the proposed solution for sex determination from human skull embodies 4 main stages: 3D data preprocessing, feature representation, dimensionality reduction, and classification as illustrated in Figure 3.1.

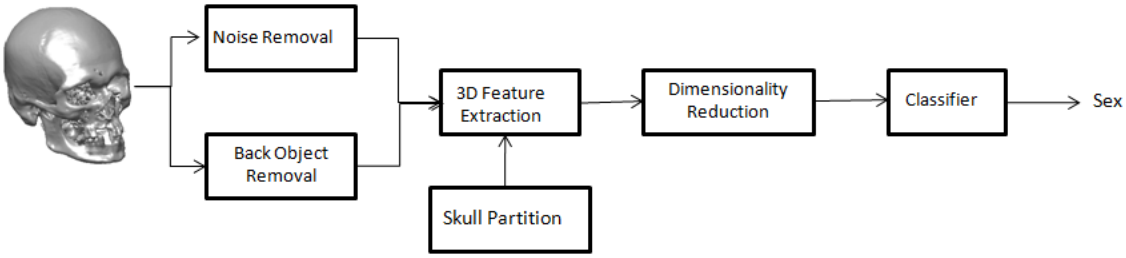


Figure 3.1: Structure of sex determination using heuristically partitioned skulls

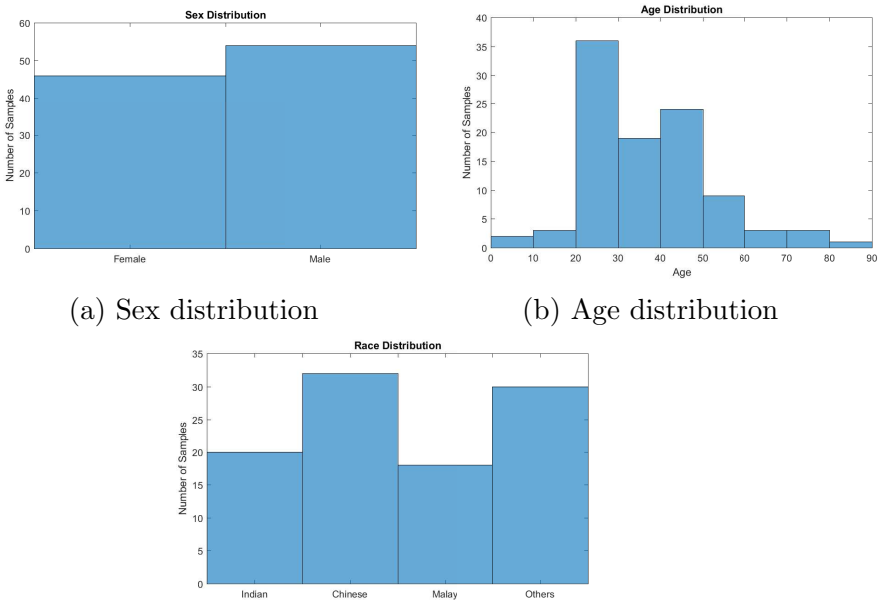
3D data preprocessing is the stage where the input volumetric PMCT sample of head region is filtered to reduce noise, remove local irregularities, and perform isosurface (mesh) reconstruction to obtain the skull layer. Feature representation involves extracting a number of 3D local shape features on the derived skull patches and the features from each local patch are learned to derive a set of compact and discriminative representations. Finally determination of sex is performed using different predictive models.

The objectives of this chapter can be summarized as:

- To describe in detail the experimental dataset, with its composition.
- To provide the detailed description of the fundamental approach proposed for sex determination with the stages, algorithms, and the choice of methods at each phase of the proposed method.

3.1.1 Experimental Dataset

The data used in this study are post-mortem computed tomography (PMCT) scans obtained from Hospital Kuala Lumpur (HKL), which have been compiled between 01-05-2012 and 30-09-2012. The demographic distribution of the data is composed of male and female subjects from south east asia (precisely 71 Malaysian and 29 Non-Malaysian). The Non-Malaysian samples are from: Bangladesh, Myanmar, Vietnam, Pakistan, Indonesia, Nepal, and Cambodia. All subjects are between the ages of 0 and 85 years. Legal consent was obtained for usage of this data prior to the commencement of the research. The 100 samples used for all experimental evaluation in this research compose of 54 male and 46 female subjects of known sex. The distribution of data is illustrated in Figure 3.2, Table 3.1, and Table 3.2.



(a) Sex distribution

(b) Age distribution

(c) Race distribution

Figure 3.2: Distribution of the 100 data samples used in the experiments in this thesis according to sex, age, and race

Table 3.1: Number of samples in different age band

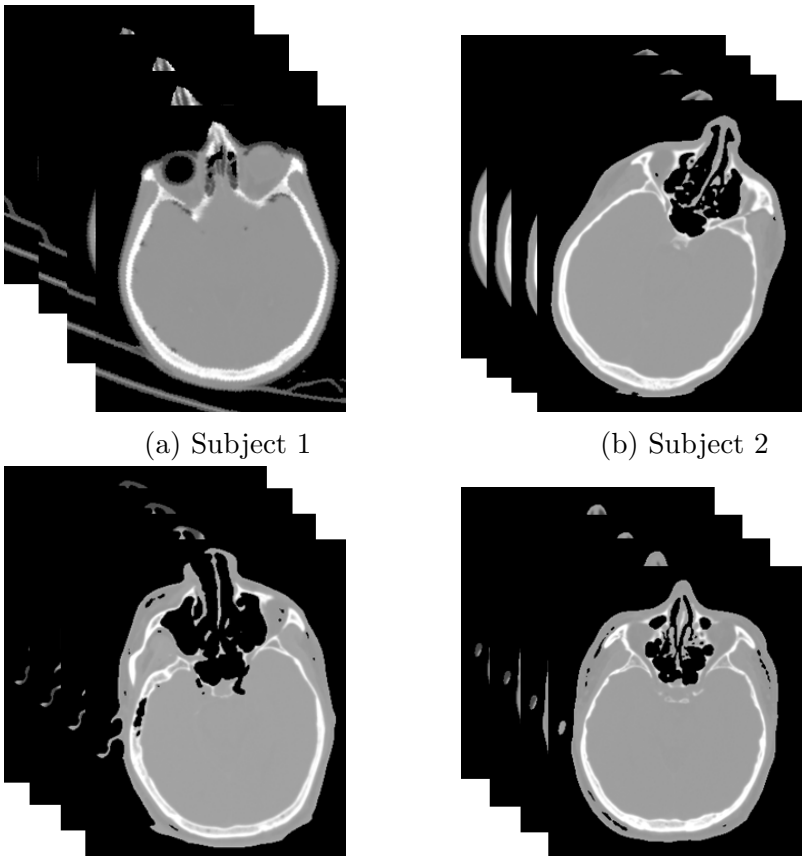
Age Range	0-25	26-45	46-65	66-85
Number of Subjects	24	53	19	4

Table 3.2: Number of samples in race groups

Race Group	Indian	Chinese	Malay	Others
Number of Subjects	M (9) F (11)	M (16) F (16)	M (15) F (3)	M (14) F (16)

NB: M - Male, F - Female, and Others - Non-Malaysian

The original format of data contains the full body of the subjects, which average between 2000 to 4000 slices per-subject and the resolution of the data is 512 $\times$ 512 $\times$ No of slices. For the purpose of this work, the head region has been segmented for each subject. The resulting input sample set is minimized to resolution between 512 $\times$ 512 $\times$ 261 and 512 $\times$ 512 $\times$ 400, depending on the size of the head region of each subject. Example of CT slices is depicted in Fig. 3.3.



(a) Subject 1 (b) Subject 2

(c) Subject 3 (d) Subject 4

Figure 3.3: Head region volume CT slices of 4 subjects prior to 3D reconstruction

3.2 3D Preprocessing

This section describes the techniques utilised to remove noise, align, and reconstruction each skull into 3D mesh from the accumulated 3D volumetric CT scans.

3.2.1 Noise Removal

Original form of 3D volumetric data is usually composed of unwanted noise and possibly background objects which should be removed to minimize the influence of outliers on prediction results. In this research, each 3D volumetric data of the head region is filtered to minimize noise, local irregularities, and roughness using a discretized spline smoothing method (Garcia, 2010). The smoothing method utilizes the 3D discrete cosine transform based penalized least square regression (DCT-PLS) on equally spaced high dimension data. PLS is a fundamental concept in thin plate splines for data interpolation and smoothing. However, according to Garcia (2010) the PLS regression problem can be reformulated with DCT, where the data are expressed in the form of cosine functions oscillating at different frequencies. The method can equally be adapted to multidimensional datasets, since DCT is multidimensional.
In the case of smoothing, the DCT-PLS method seeks to minimize the following Garcia (2010):

$$\begin{aligned} S(x') &= RSS + sP(x') \\ &= ||x' - x||^2 + \ sP(x') \end{aligned} \tag{3.1}$$

where, x is the noisy data, x' is smooth data, RSS is the residual sum of square error, equivalent to the euclidean norm $||x' - x||^2$, s is a small scalar value that determines the smoothing rate (a larger s leads to smoother x'), and P is a penalty term that expresses the roughness of the smooth data. When P is expressed as the square integral of the pth derivative of x' then PLS regression is simply a smoothing spline. However, according to Garcia (2010) an alternative method to express the roughness of the smooth data can be done using a second-order divided difference:

$$P(x') = \ ||Dx'||^2 \tag{3.2}$$

where D is a tridiagonal square matrix. Hence, using Eqs. (3.1) and (3.2) $S(x')$ can be minimized using the following system of linear equation to obtain the smoothed data:

$$(I_n + sD^\top D)x = x' \tag{3.3}$$

where I_n is a unit matrix and $D^\top$ is the transpose of D. The best value of s can be determined by generalized cross-validation (GCV). To apply the method to an evenly spaced gridded data, D can be rewritten as a simple difference matrix (Garcia, 2010):

$$D = \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix}.$$

Then the eigendecompsition of D is

$$D = U \Lambda U^{-1} \tag{3.4}$$

where Λ is the eigenvalues of D and since U is a unitary matrix, it follows the relation:

$$x' = \; U \left(I_n + s\Lambda^2\right)^{-1} U^\top x \;\equiv\; U \, \Gamma \, U^\top x \tag{3.5}$$

where Γ is a diagonal matrix and because U and $U^\top$ are equivalent to n-by-n type-2 DCT and inverse DCT, so the smooth output x' can be further expressed as:

$$x' = U \, \Gamma \, \mathrm{DCT}(x) = \mathrm{IDCT}\left(\Gamma \mathrm{DCT}(x)\right) \tag{3.6}$$

Now, to deal with multi-dimensional data, the above derivation can be easily extended by:

$$x' = \mathrm{IDCTN}\left(\Gamma^N \circ \mathrm{DCTN}(x)\right) \tag{3.7}$$

where DCTN and IDCTN are the N-dimensional discrete cosine transforms, $\circ$ represent the Schur (elementwise) product, and Γ^N is a tensor of rank N defined as:

$$\Gamma^N = 1^N \div \left(1^N + s\Lambda^N \circ \Lambda^N\right) \tag{3.8}$$

where $\div$ is element-by-element division and 1^N is an N-rank tensor of ones. After noise removal, the marching cubes algorithm is adopted to reconstruct the volume data into a 3D surface (Lorensen and Cline, 1987), as shown in Fig. 3.4.

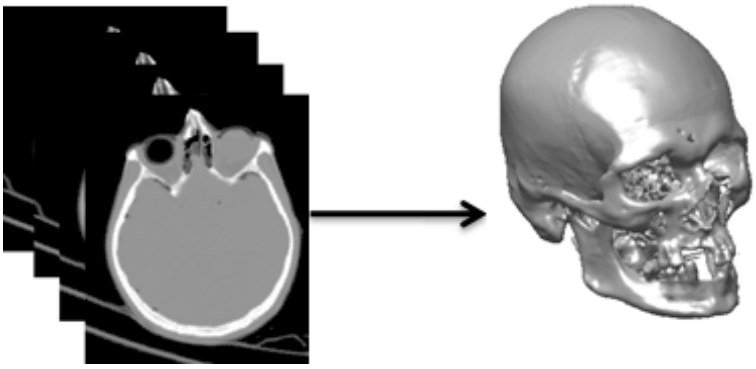
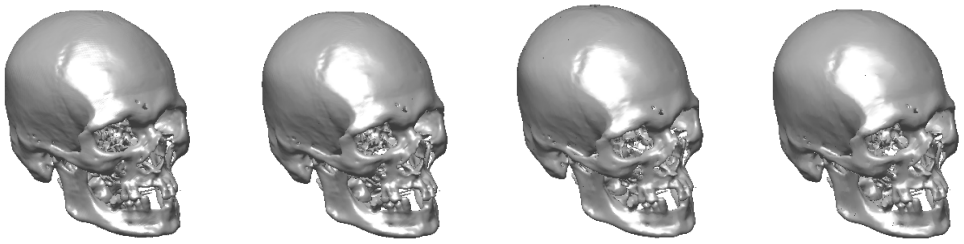


Figure 3.4: 3D reconstruction from volume slices using Marching cubes algorithm

In this thesis, the reconstruction is achieved using an iso-value of (150) to obtain the regions containing the hard tissues. Due to the high dimension of the reconstructed skulls (for instance, the number of vertices in the original reconstructed skull is between 650000 and 850000 vertices), it is preferable to perform downsampling in order to reduce computational complexity and time. As a result, mesh simplification (Cignoni et al., 1998) is performed on the original reconstructed mesh to downsample the 3D surface to 13% of the original size. The resulting downsampled skulls have < 130000 vertices, but the structural details of the surface are still well preserved after downsampling, as it can be seen in Fig. 3.5.



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

Figure 3.5: Examples of Reconstructed Skull data. (a) Original Skull (b) Downsampled to 50% (b) Downsampled to 25% (d) Downsampled to 13%.

3.2.2 Background Object Removal

Considering the fact that the data samples are real data obtained from hospital, some dynamic background objects were noticed, which have the same isovalue as the hard tissue, on some skull samples (approximately 27 samples) after surface reconstruction. To remove these unwanted or noisy background objects, this thesis proposes to learn an online sequential linear systems with gaussian mixture model sample initialization. Firstly, 8 samples of clean skulls and 8 samples of noisy

skull data are randomly selected. Manual annotation is perform on the noisy skull data to cut out the noisy objects. The set of segmented noisy objects are used to form the negative training samples, while the clean skulls are regarded as the positive training samples. Therefore, the training set contains 8 clean skulls and 8 segmented noisy object regions resulting in $\sim$ 750000 vertex points for the two classes. Secondly, due to the size of the training set, a sequential learning method is utilized, where each time the training is performed in batches. However, the first batch of data for training is initialized by using GMM to fit a mixture model composing of n components. Given a training set $A = \{v_1, \dots v_i\}$, where A is a matrix composing of clean and segmented noisy objects, and v_i is a vertex point, the formulation of the n components GMM can be expressed as:

$$p(v|\lambda) = \sum_{i=1}^{n} w_j \; g(v|\mu_j, \Sigma_j) \tag{3.9}$$

where μ_j is the mean vector, Σ_j is the covariance matrix, w_j are the weights of the mixture and $g(v|\mu_j, \Sigma_j)$ are the gaussian density of the components. The entire parameters can be placed as $\lambda = \{w_j, \mu_j, \Sigma_j\} \;\; j = 1, \dots n$, which need to be estimated. There are several methods that can be used for parameter estimation, but the one used in this research is the expectation maximization (EM) algorithm. It is an algorithm which solves two main steps: the expectation **E-step** and the maximization **M-step** with an initial model λ to estimate a new model $\bar{\lambda}$, until the expected value converges or reaches the maximum number of iteration.

Thus, the sample set A is clustered into n-clusters with mixture of gaussians. To initialize the training process of background object removal, a cluster from output of GMM is selected as the initial batch A_0. After selecting the initial batch, the following minimizing problem of a linear system is solved as:

$$\text{Minimize: } ||A_0 x_0 - b_0||^2 + \lambda ||x_0||^2 \tag{3.10}$$

where x_0 is the learned coefficients with the smallest norm, λ is a regularization value, and $b_0 = \{+1, -1\}$ are the labels (+1 for clean skull, -1 for noisy object) of the initial subsets. The initial solution x_0 can be obtained analytically using:

$$x_0 = \left(A_0^\top A_0 + \lambda \mathrm{I}\right)^{-1} A_0^\top b_0 \tag{3.11}$$

Formally, the coefficient x_0 is regarded as the least square solution to the minimization problem. The inverse matrix M_0^{-1} can be obtained as: $M_0^{-1} = \left(A_0^\top A_0 + \lambda \mathrm{I}\right)^{-1}$ and the output coefficient $x_0 = M_0^{-1} A_0^\top b_0$.

Updating x_{k+1}

Now, the interest is to train the remaining subset of data in batches also, but without keeping the previously trained set in order to reduce computational time and complexity. Therefore, it is preferable to find an efficient approach to update the initially learned coefficient, while solving new linear systems minimization problem (Do and Fekete, 2007). Assuming the least square solution to the minimization problem previously initialized with GMM clustering is obtained in the form of x_0, one well recognized method to subsequently update x_0 is the sherman-morrison woodbury inverse formula (Golub and Van Loan, 1996). When new sample subsets $A_{k+1} = \{v_1^{k+1}, \ldots, v_i^{k+1}\}$ are learned in batches x_0 can be updated from Eq. 3.11 as:

$$x_{k+1} = x_k + M_{k+1}^{-1} A_{k+1}^T \left(b_{k+1} - A_{k+1} x_k\right) \tag{3.12}$$

$$M_{k+1}^{-1} = M_k^{-1} - M_k^{-1} A_{k+1} \left(\mathrm{I} + A_{k+1}^T M_k^{-1} A_{k+1}\right)^{-1} A_{k+1}^T M_k^{-1} \tag{3.13}$$

where x_k is the output coefficient in the previous step, b_{k+1} is the newly collected labels and $M_{k+1}^{-1} = \left(A_{k+1}^T A_{k+1} + \lambda \mathrm{I}\right)^{-1}$. Once x_{k+1} has been learned, each new skull sample containing noise, which were not included in the training phase are used to test the model. The results obtained are depicted in Fig. 3.6.

3.2.3 Mesh Alignment

Mesh alignment is a necessary stage for centralizing the 3D coordinates of mesh points. In order to centralize the pose of the subjects, mesh alignment is performed using iterative closest point (ICP) method (Besl and McKay, 1992). ICP normally requires a set of predefined interest points from a model and target mesh to iteratively perform the alignment by minimizing the sum of square error between the two data. The approach in this thesis is to generate the interest points by using the difference of gaussian (DoG) scale space technique, computed on a shape primitive (first order information).

Considering a set of vertices $V = \{v_1, \ldots, v_n\}$ points from a triangulated mesh with N neighborhood connectivity defined on V. The mesh is filtered with gaussian filter using different smoothing parameters σ_s, which increase exponentially by $\sigma_s s = 2^s$ (this work uses 20 scales). Then, gaussian curvature (a shape primitive function) $G_c = k_{max} \cdot k_{min}$ at each scale is computed, where k_{max} and k_{min} are the maximum and minimum curvature of a mesh point.

Afterward, the difference between the gaussian curvature of the smoothed mesh obtained at different scales is computed. Finally, the vertices with the highest DoG

values in a neighbourhood (local maxima) are selected as the detected interest points. Once the interest points have been detected, a sample mesh is selected as the model and every other mesh data is subsequently registered to the model using ICP to obtain the transformation parameters (rotation matrix and translation vector). These parameters are used to transform the target data. Some aligned skulls are depicted in Fig. 3.6.

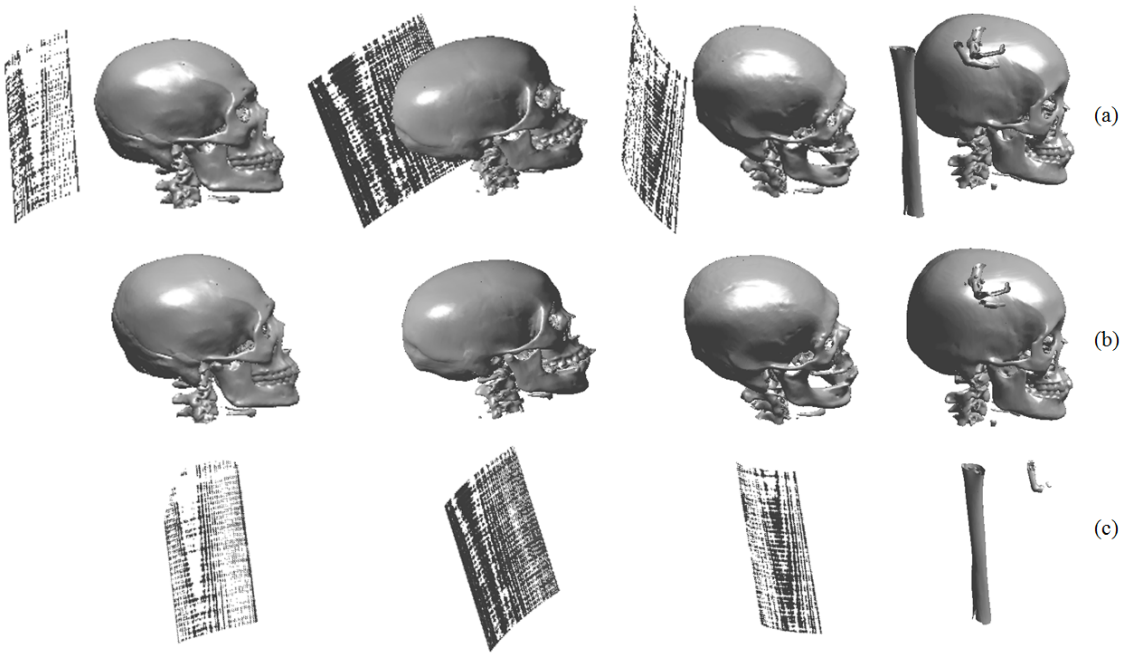


Figure 3.6: Noisy objects removal (a) data with noise (b) clean and aligned skulls after object removal (c) the removed noise

The registered data are used in the next stage for feature extraction with different local descriptors.

3.3 3D Local Shape Descriptors

The representative features computed from the registered mesh data are mesh local binary pattern (MeshLBP), Spin Image, Local Depth SIFT (LD-SIFT), scale invariant heat kernel signature (SIHKS), ShapeDNA, and Region Area. Out of the 6 techniques only MeshLBP, SIHKS, ShapeDNA, and Region Area are extracted on the entire skull because the other 2 methods are keypoint based. The main reason for considering these 6 techniques is because they encapsulate the characteristics or properties of multitude of local descriptors and their variants reported in the literature.

As a general practice in computer vision, when features are extracted locally, feature concatenation can be adopted to transform the representations into a feature vector, since predictive models theoretically perform vector operations. However, if local representation from each point are concatenated, the resulting features will be in the order of 10^5, which is a high dimension. As such classifiers may find difficulty in separating such huge features because of computational complexity in addition to the high computational time. With regard to this, different techniques are used to reduce the features into compact representation.

3.3.1 Mesh Local Binary Pattern (meshLBP)

MeshLBP is based on extracting a set of ordered ring facets (ORF) (Werghi et al., 2012) and calculating a primitive function such as mean or gaussian curvature for each facet in the ring. Then the primitive function of centre facet is used as a thresholding function to obtain the binary patterns of the neighbouring facets (Werghi et al., 2015a). Consider a primitive function $p(f)$ defined on the mesh, which represent the gaussian curvature. The mesh-LBP can be computed as follows:

$$meshLBP^{r}_{m}(f_c) = \sum_{k=0}^{m-1} s(p(f^r_k) - p(f_c)) \cdot \alpha(k) \tag{3.14}$$

$$s(x) = \begin{cases} 1, & > 0 \\ 0, & \leq 0 \end{cases}$$

where r is the ring number, and m is the number of facets in the ring.

The two parameters control the radial resolution and azimuthal quantization. The added parameter $\alpha(k)$ is a discrete function that allows the computation of other LBP variants. To extract MeshLBP in this work, 10-ring neighbours are computed for each unique facet. Then the MeshLBP is extracted by comparing the gaussian curvature of centre facet to the neighbouring facets. An example of MeshLBP feature representation is illustrated in Figure 3.7. The output of MeshLBP is V-by-10 matrix for each sample data, with V denoting the number of vertex points. It takes $\sim$ 1325 sec to extract MeshLBP features for a sample cranial data.

3.3.2 Spin Image

It is a state-of-the-art descriptor proposed by Johnson (1997) for object retrieval. The descriptor is invariant to rotation and translation. The main idea of spin image is to generate a 2D histogram for mesh point composing of representation of

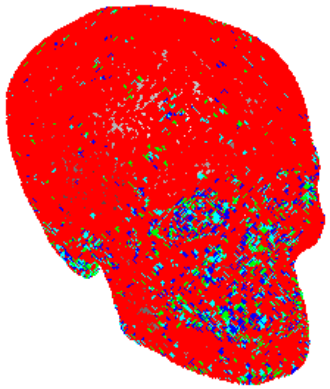


Figure 3.7: Example of MeshLBP feature distribution by mapping the binary patterns to different color map. For illustration, the binary patterns of ring (r = 2) are displayed. Red color indicates areas of small binary pattern (0) due to small variation in the curvature of the local regions. Blue represents patterns with values (1 16), green denotes patterns with values (32 256), and cyan represents patterns with values (512 4096), due to large variation in the curvature of the local regions. For instance, sharp curvature variation can be observed around the local regions of the orbital, nasal bone, and mandible, while the parietal bone exhibits mild variation in curvature

the object geometry. In essence, it performs representation of object's coordinate system rather than viewer's coordinate system. Oriented points are computed for each point on the mesh based on the position of the vertex p and surface normal vector n. Based on this, a 2D basis (p, n) is formulated. Then to obtain the coordinate system, a tangent plane P passing through p is formed which is perpendicular to the normal n and line L through p that is parallel to n. This computation results into a cylindrical coordinate system (α, β), with α being a non-negative distance to L and β a signed distance perpendicular to P.
These oriented points are used to construct a spin map S_o, which is simply a projection of 3D points x on a mesh to 2D coordinates (α, β). The mapping can be expressed as:

$$S_o : R_3 \rightarrow R_2$$

$$S_o(x) \rightarrow (\alpha, \beta) = \left(\sqrt{||x-p||^2 - (n \cdot (x-p)^2}, n \cdot (x-p)\right) \tag{3.15}$$

The 2D points are further accumulated into discrete points which are incrementally updated for each new 2D points. The size (i_{max}, j_{max}) of a spin image is determined by the size of the bin and the maximum size of the object expressed in the spin map coordinates:

$$i_{max} = \frac{2\beta_{max}}{b} + 1 \ , \qquad j_{max} = \frac{\alpha_{max}}{b} + 1 \tag{3.16}$$

The spin image is extracted over a 10×10 grid positioned over the keypoint. The spin image for a cranial data can be extracted in ~ 125 sec.

3.3.3 Local Depth SIFT (LD-SIFT)

LD-SIFT is an extension of the original 2D based SIFT to mesh surface, presented by Darom and Keller (2012). Given an interest point that is a local maxima resulting from difference of gaussians (DoG) operation on the mesh, a sphere support is constructed around it covering a neighbouring region. LD-SIFT estimates a tangent plane T within the support region, and compute the distance from each point in that region to the plane to create a 2D array of depth map. Then, it sets the viewpoint size to match the feature scale, as detected by the DoG detector, which enables the construction to be scale invariant. Further, it computes the principal component analysis (PCA) of the the points surrounding the interest point, and use their dominant direction as the local dominant angle.
Similar to the standard SIFT by Lowe (2004), the depth map are rotated to a canonical angle based on the dominant angle, which makes the LD-SIFT rotation invariant too. This mainly differentiate LD-SIFT from Spin Image, as the angles in Spin Image are obtained based on the normal vectors. From the resulting depth maps, the standard SIFT feature descriptors are computed to create the LD-SIFT feature descriptor. The features are mainly representation of 8-bins gradient histograms distributed in the local cells of 4-by-4 depth map. The LD-SIFT for a cranial data can be extracted in ~ 1085 sec.

3.3.4 Scale Invariant Heat Kernel Signature (SIHKS)

Heat kernel signature (HKS) is a spectral shape representation that provides the point signature of a specific point in a mesh via deformable shape analysis (Sun et al., 2009). The concept is based on modelling shapes as Riemannian manifold and use their heat conduction properties as descriptor. The heat propagation can be generalized by the Heat diffusion equation:

$$\left(\triangle_X + \frac{\partial}{\partial t} \right) u = 0 \tag{3.17}$$

where $\triangle_X$ is the Laplace-Beltrami operator and the solution $u(x,t)$ of the heat diffusion describes the amount of heat on a surface at a particular time t. The solution of Eq.3.17 with point heat distribution $u_0(x) = \partial(x - z)$ as initial conditions is called the heat kernel and denoted by $K_{X,t}(x,z)$. The equation can

be compactly expressed as:

$$K_{X,t}(x,z) = \sum_{i=0}^{\infty} e^{-\lambda_i t}\phi_i(x)\ \phi_i(z) \tag{3.18}$$

where λ_i are eigenvalues and ϕ_i are eigenfunctions.

The original heat kernel signature for shape representation was proposed by Sun et al. (2009), which is invariant to isometry and its has been popularly used for shape retrieval in computer vision applications. However HKS has the limitation of scale invariance, thus SIHKS was recently introduced to remove scale effect by sampling each point logarithimically in time ($t = \alpha^{\tau}$) (Bronstein et al., 2010). Then calculating the derivative with respect to the scale to undo the additive function with respect to the scale. Further, they used fourier transform to cancel out the shift variation. To extract SIHKS, a logarithmic scale-space with base $\alpha = 2$ ranging from $\tau = 1 : 20$ with increments of $1/5$ is utilized in this research. The first 10 frequencies are then selected for local description. Each of the feature point in the 10 length vector is the heat kernel signature at a particular scale. In other words, each column contains the frequencies of HKS of the whole mesh at that scale. It takes ~ 58 sec to extract the SIHKS descriptor for a sample cranial data.

3.3.5 ShapeDNA

ShapeDNA initially proposed by Reuter et al. (2006), computes the fingerprint of any 3D object by deriving the eigenvalues of laplace-beltrami operator of the shape model. It is isometry invariant and insensitive to noise. Normally, eigenvalues λ and eigenfunctions u are the solutions to the laplacian eigenvalue problem $\triangle u = -\lambda u$, where $\triangle u := div(grad(u))$, div is the divergence of the underlying Riemanian manifold and $grad$ represents the gradient. The first normed N eignevalues $0 < \lambda_1 < \lambda_2, \ldots \lambda_n$ are chosen as the shape descriptor. In this report, the first 100 eigenvalues are selected as the shape descriptor. Figure 3.8 depicts some examples of ShapeDNA representation.

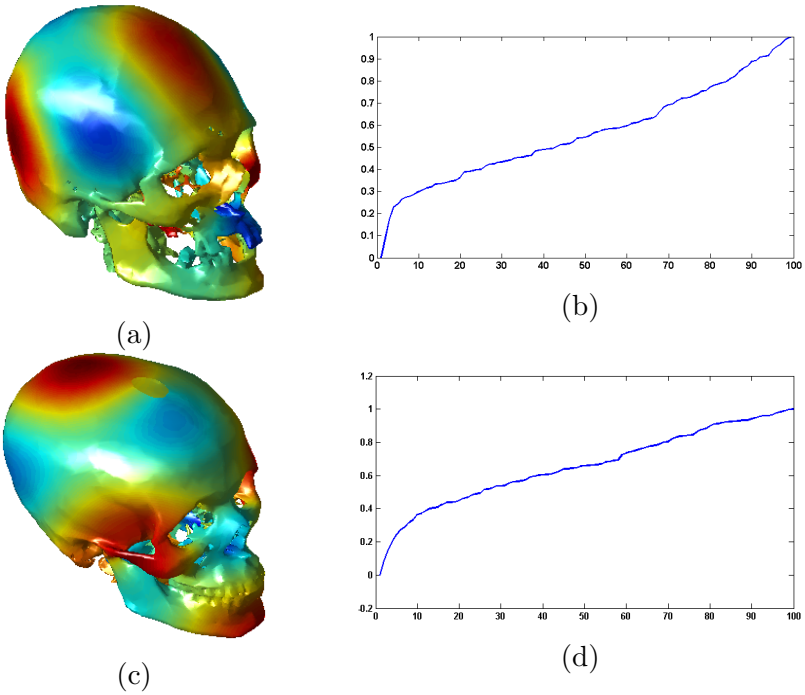


Figure 3.8: (a) and (c) Eigenvector over the rendered 3D skull surface. The areas with small eigenvectors are represented with color blue and as the values increase, the color goes from green to red (b) and (d) Plot of 100 eigenvalues forming ShapeDNA descriptor.

3.3.6 Area of Skull Regions

Finally, the surface area of each local sub-region of skull is computed as additional features. The skull is initially partitioned in smaller regions along the X,Y, or Z and the corresponding area is computed following the approach in (Zhang and Chen, 2001).

3.4 Feature Learning

To further obtain comprehensive representation from the extracted features, two dimensionality reduction methods are adopted: Kernel Principal Component Analysis (KPCA)(Schölkopf et al., 1998) and a dictionary learning technique: K-singular value decomposition (K-SVD) (Aharon et al., 2006).

3.4.1 KPCA

KPCA is a non-linear modification to the standard PCA, which is commonly used for dimensionality reduction. The essence of using KPCA is to transform the features from their current space where they are well correlated to a lower dimensional space.

KPCA can be implemented in few steps:

- Select a proper kernel function linear, radial basis function (RBF) and derive the kernel matrix

$$K = k(x, x_i) = (\phi(x), \phi(x_i)) = \phi(x)^T \phi(x_i) \tag{3.19}$$

- Compute the Mean and Covariance

$$\text{Mean} : \frac{1}{N} \sum_{i=1}^{N} \phi(x) = 0 \; ; \quad \text{Cov} : \frac{1}{N} \sum_{i=1}^{N} \phi(x)^T \phi(x_i) \tag{3.20}$$

- Solve for the eigenvectors

$$\text{Eigenvectors} : Cv = \lambda v \tag{3.21}$$

3.4.2 K-SVD

KSVD is a dictionary learning algorithm, which can be implemented as follows:

Find an overcomplete dictionary matrix $D \in \mathbb{R}^{n \times k}$ that contains K prototype signal-atoms for columns.

$$\min_{D,W} ||X - DW||_F^2 \quad s.t \quad \forall i \;\; ||w_i||_0 \leq T \tag{3.22}$$

where X contains the sample data, W is the coefficient matrix, and T is a threshold value.

- Iteratively minimize Eq. 3.22 by first fixing D and finding the coefficient matrix W using any pursuit (optimization) method. Then search for a better dictionary.

- Update one column at a time, fixing all the other columns, and find a new column d_j and new values for its coefficients that best minimize the mean square error (MSE).

- The solution of updating only one column of D at a time can be directly obtained through singular value decomposition (SVD).

3.5 Classification

In order to perform classification, support vector machine (SVM) (Cortes and Vapnik, 1995), kernel extreme learning machine (K-ELM) (Huang, 2014), sparse representation classifier (SRC) (Wright et al., 2009) are adopted in this thesis.

3.5.1 Support Vector Predictive Models

Support vector machine (SVM) is a structural risk minimization algorithm based on statistical learning theory. The main idea of SVM is aimed at finding an optimal separating hyperplane that sufficiently separates the data. Detailed mathematical theorems on SVM can be found in (Cortes and Vapnik, 1995). In this thesis, the soft margin SVM, usually considered when the input data space is not linearly separable is adopted. In this case, the soft margin hyperplane can be found by incurring slack variables ξ_i derived from allowing some errors during training (Cortes and Vapnik, 1995). The problem of finding the hyperplane with minimizing the training errors can be expressed as follows:

$$\begin{aligned} \text{Min: } & \frac{1}{2}w^\top w + C\sum_{i=1}^{M}\xi_i \\ \text{s.t. } & y_i(w(x_i)+b) \geq 1-\xi_i, \quad i=1,\ldots,M \end{aligned} \tag{3.23}$$

where w is the normal to the optimal separating hyperplane and b is the bias term. The user defined parameter C in the equation is the cost parameter which determines the tradeoff between the training error and distance of hyperplane. The standard SVM is a binary classifier which can also be extended to multi-classification via 1-v-1 or 1-v-all winner takes all techniques. However, it cannot be used for regression analysis directly, which the reason why its variant support vector regression (SVR) has been introduced.

3.5.2 Kernel Extreme Learning Machine (K-ELM)

ELM is a single layer feed forward neural network, with randomly generated hidden neurons, with the output layer solved analytically. ELM can be used for classification, regression, and clustering. The output function of ELM can be written as (Huang et al., 2006):

$$f(x) = \sum_{j=1}^{M} Q_j K(x_n c_j + d_j) \qquad n = 1,2,\ldots,N \tag{3.24}$$

where $c_j \in R$ is the randomly generated bias of the jth hidden node and $d_j \in R$ is the randomly generated input weight vector connecting the jth hidden node and input nodes. Q_j is the weight vector connecting the jth hidden node to the output node. $K(x_n c_j + d_j)$ is the output of the jth hidden node with respect to the input

sample x_n. Its generalized form can be express as (Huang et al., 2006):

$$E\beta = T \tag{3.25}$$

and the output weight β can be obtained by solving the least square of:

$$\beta = \left(\frac{\mathrm{I}}{C} + EE^{\top}\right)^{-1} E^{\top}T \tag{3.26}$$

where E is the hidden layer matrix, C is the regularization factor, and T is the target value.
There are three main steps involved in ELM:

· Randomly assign input weight Q_j and bias c_j.

· Calculate the hidden layer output matrix E

· Calculate the output weight β by $\beta = E^{\dagger}T$.

where $E^{\dagger}$ is the Moore-Penrose generalized inverse of E. ELM can be extended to the non-linear classification problem by using a kernel function to map the input feature space (Huang, 2014). Thus the E matrix that contains the randomly generated neurons can be replaced with a kernel matrix obtained using a transformation function such as linear, RBF or polynomial.

3.5.3 Sparse Representation Classifier

SRC is based on the concept of compressing sensing and low rank approximation. The idea is quite straightforward, which is to represent a test sample as a sparse linear combination of the entire training samples. Intuitively, the sparse coefficients (non-zeros elements) would denote the training samples that belong to the same class as the test sample. Much like ELM but without neurons, SRC also depends on a linear equation $z = Ax$ where $A \in R^{N\times M}$ training matrix. However, the problem of SRC is generally formulated as an underdetermined system of linear equation, thus $N < M$ in the training matrix. SRC was first introduced by Wright et al. (2009) for face recognition, and since then it has gained wide application in object recognition problem in the domain of computer vision.

Assuming there are c classes of patterns. Let A_i be the matrix of training samples of class i. In other words $A_i = [y_{i1}, y_{i2}, \ldots, y_{iM_i}] \in R^{d\times M_i}$ where M_i is the number of training samples of class i. Let us define a matrix $A = [A_1, A_2, \ldots, A_c] \in R^{d\times M}$, where $M = \sum_{i=1}^{c} M_i$, so A is a matrix containing all training samples. Given a test sample z, it can be represented as a linear combination of an overcomplete

dictionary and sparse coefficients x whose basis lie in the column space of matrix A i.e $z = Ax$. The sparse vector x can be obtained via L_1 minimization:

$$x_1 = \text{argmin} \, ||x||_1 \ , \quad s.t. \ \ Ax = z \tag{3.27}$$

There are several linear programming algorithms to solve the minimization problem in polynomial time. Once the sparse solution has been obtained, it will be noticed that the vector contains few non-zero elements. Thus the classification can be performed by assigning z to the class with the largest value in x. The corresponding index to the coefficient indicates the class of z.

3.5.4 Multiple Kernel Learning (MKL)

Besides testing the performance ability of the 6 local shape descriptors independently, the potential of MKL to fuse multiple features with different representation properties is examined. MKL has attracted significant amount of attention in CV research domain. In this thesis, the soft margin MKL algorithm introduced by Xu et al. (2013) is adopted, where a kernel slack variable is first introduced for each of the base kernels when learning the kernel. This approach is advancement over the MKL algorithm generally regarded as the hard margin MKL (Gönen and Alpaydn, 2011), which imposes sparsity on a category of features and select the features that best optimize the object function. The motivation behind MKL is to combine or fuse multiple kernels or features rather than using a single feature representation to make prediction, with expectation that such combination lead to potential gain in performance. Assuming, there is a training set M consisting of $(h_i, g_i, w_i)_{i=1}^{M}$ where h_i is a type of local features of the ith sample and g_i is a another type of features of the same sample, and $w = (-1, +1)$ are the class labels. The features can be mapped to the reproducing kernel hilbert space (RKHS) with a kernel function $k(\cdot, \cdot)$ defined over each of the two feature types. Without loss of generality, this notion applies to more than 2 feature types.

SVM based MKL

Suppose, there is a set of base kernels $K = \{K_1, \dots K_N\}$ for the training set, the objective of the standard problem solved by the MKL introduced in (Xu et al., 2013) is:

$$\max_{\alpha \in \mathcal{A}} \min_{\mu \in \mathcal{M}} \sum_{n=1}^{N} \mu_n \mathbf{SVM}\{K_n, \alpha\} \tag{3.28}$$

where, $\mathbf{SVM}\{K_n, \alpha\} = -1/2(\alpha \odot w)' K_n(\alpha \odot w)$, α_i are the coefficients of the samples, and $\mu = [\mu_1, \ldots \mu_N]'$ [1] are the coefficients measuring the importance of the nth base kernel. $\mathcal{A} = \{\alpha | \alpha' \mathbf{1} = \mathbf{1}, \alpha' \mathbf{w} = \mathbf{0}, \mathbf{0} \leq \alpha \leq \mathbf{C}\}$ is the domain for α and $\mathcal{M} = \{\alpha | \mathbf{0} \leq \mu, \sum_{\mathbf{n=1}}^{\mathbf{N}} \mu = \mathbf{1}\}$ is the domain for μ. For the solution in Eq. 6 a hard margin is constructed resulting in selection of only the most important base kernel, which eventually defeat the purpose of finding interaction between different feature set. Therefore, in this thesis, the soft margin formulation is used, where a slack variable which is the difference between the target margin τ and the objective, is introduced as:

$$\zeta = \tau - \mathbf{SVM}\{K_n, \alpha\} \quad \forall_n = 1, \ldots N \tag{3.29}$$

The loss incurred by the kernel slack can be expressed as

$$\xi_n = \ell(\zeta_n) \quad \forall_n = 1, \ldots N \tag{3.30}$$

where $\ell(\cdot)$ is a hinge loss function $\ell(\cdot) = max(0, \ell_n)$.

[1] the notation $'$ indicates a transpose

Chapter 4

An Empirical Study on the Typical 3D Shape Descriptors and their Patch Representations for Sex Identification

> The best journey of life that you can take is inside yourself to discover your new thoughts, new ideas, and new hopes
>
> Debasish Mridha

This chapter presents the experimental attempts on sex determination from skull data. It provides the detailed description of the methods that have been explored for pre-processing of the skull dataset, feature extraction, feature learning, and classification techniques adopted for prediction of demographic characteristics.

4.1 Overview

The approach introduced in this chapter is composed of 4 main stages: 3D data preprocessing, skull partitioning, feature extraction, and classification as illustrated in Figure 4.1.

The 3D preprocessing is the stage where the input volumetric PMCT sample of head region is filtered to reduce noise and smoothen to remove irregularities, then isosurface (mesh) reconstruction is performed to obtain the skull layer. The stage of skull partitioning involve dividing the skull into different regions locally along the X, Y and Z axes, representative of the anatomical axes: axial, coronal, and sagittal. The feature extraction involves extracting a number of local features on the reconstructed mesh and the features are learned to derive a set compact and discriminative representations. Finally determination of sex is performed using different predictive models.

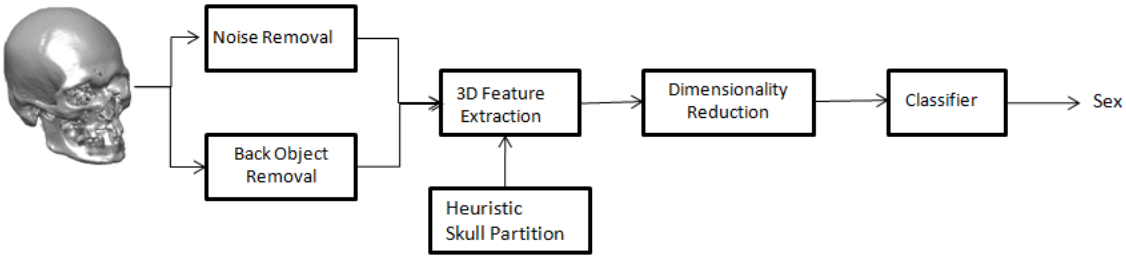


Figure 4.1: Structure of sex determination using heuristically partitioned skulls

The focus of this chapter can be summarized as:

- Examine the performance of different 3D local shape descriptors on sex determination.
- Assess the effect of orientation of partition and number of skull partitions on sex determination.

4.2 Patch based Representation and Feature Aggregation from Heuristically Partitioned Skulls

One of the first main challenges faced in this research is the decision on how features extracted on the skull can be represented to offer meaningful and distinctive interpretation of the skull.

Based on established profound concepts in computer vision, once features have been extracted on the entire or substantial keypoints subset of a sample data, the popular techniques that could be considered is to either concatenate the keypoints into a long-tailed feature vector or form compact representation with bag of words (BoW) model. However, from the initial experiments on the skull data depicted in Figure 4.6 Section 4.3.2, it has been empirically discovered that these techniques generally result in a high possibility of features from a particular class having dissimilar representation and as a direct consequence, features from two different classes may exhibit the same or similar values which reduced the informative characteristics of such methods.

Therefore, to minimize such inseparability, in this thesis the skull is divided into different regions. Due to the fact that, there is no information about the location of the anatomical sites of the skull, the skull is partitioned with equal spacing along the geometric axes, as this geometric constraint could guarantee proper region distributions. Then, the entire features from the vertices in each region are aggregated into a single descriptor by stacking those representations.

For instance, assuming spin image of 10×10 size is extracted from $V = 10000$ vertices, the spin image descriptors are stacked from each vertex point on one another and aggregated, which results in a final spin image of 10×10 descriptor size and feature vector of 100 dimensions, as shown in Fig. 4.2. This makes the representation more distinctive than the long-tailed concatenated or bag of word representation. Similar approach is used for LD-SIFT by stacking the gradient features from the keypoints and taking aggregate, which results in a final feature vector of 128-dimension.

To make the evaluation of the extracted 3D local descriptors comprehensive, the partitioning of the skull is performed along the X, Y, and Z axes representative of the coronal, axial, and sagittal axes, as depicted in Fig 4.4 and the number of regions, N examined in each axis is $N = \{1 : 2 : 99\}$. The results attained from the regions in each axis are illustrated in section 4.3.3.

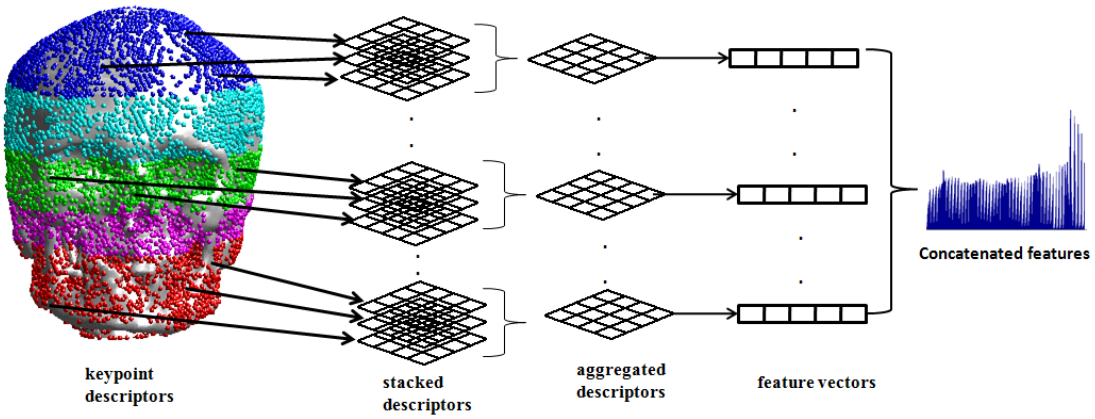
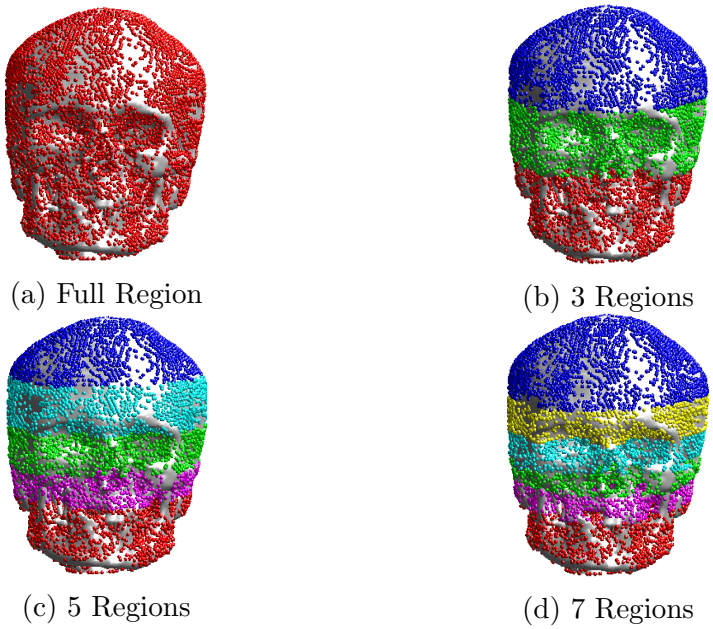


Figure 4.2: An example of multi-region representation with LD-SIFT. Each color denotes a partitioned sub-region of the skull.

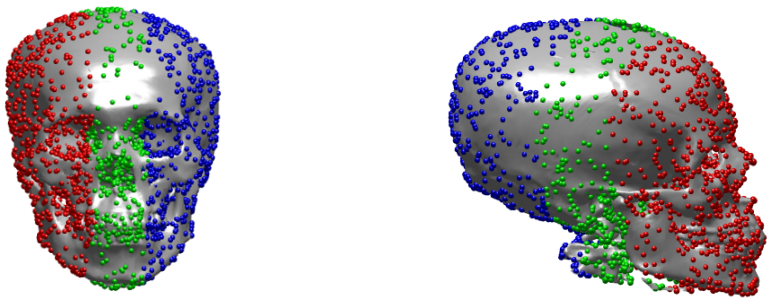
For MeshLBP, the descriptor extracted is 10 scalar values from 10 rings for each vertex point, making a $V \times 10$ descriptor matrix, thus a 32 histogram bin is aggregated for each ring along the vertices and the final feature vector is the concatenation of the histograms from the 10 rings ($32 \times 10 = 320$ feature descriptor). 32 bins is used to keep the dimension of the feature vector to a reasonable length. Also, SIHKS is compactly represented in this fashion, using 32 bins histogram bins to represent each frequency, which yields $32 \times 10 = 320$ descriptor. Examples of region partitioning of different sizes are shown in Fig 4.3. The results attained using the skull as a whole and multi-region partitioning are reported in Section 4.3.



(a) Full Region (b) 3 Regions

(c) 5 Regions (d) 7 Regions

Figure 4.3: Illustration of multi-region feature representations using different number of partitions. (b)-(d) show skull partitioning performed along the Z-axis on skull sample 1



(a) 3 Regions (X-axis) (b) 3 Regions (Y-axis)

Figure 4.4: Some examples of multi-region partitioned skulls along different axes. (a) on skull sample 7 (b) on skull sample 4

4.3 Sex Determination using Patch based Representations from Heuristically Partitioned Skulls

This section presents the experimental results and performance comparison of the described feature extraction and classification techniques. In each of the demographic estimation experiments, the input dataset is divided into 60% and 40% as training and testing set respectively by random sampling without replacement. During training a 5-fold cross validation is performed in order to determine the best model regularization value C (between 2^{-10} and 2^{10}) for SVM, SRC, and K-ELM. After training is completed, the demographic estimation is achieved by testing the

model with the separate unused testing set. Each experiment has been repeated 10 times, with the average of 10 runs calculated as the final estimation result.

4.3.1 Initial Analysis using BoW representation

Initially, the evaluation is performed to investigate the sex determination accuracy obtainable when the 3D local descriptors are transformed in BoW representation. As the BoW has commonly been used with SIFT, Spin Image, and HKS, therefore the technique is adopted to form compact description of those descriptors. To use the BoW compact representation, 35 samples from the training set of nearly equal number of males and females are randomly chosen to build a codebook (or dictionary). The codebook is generated by sampling 3000 points from each cranial for k-means clustering. For instance, with LD-SIFT the selected subset points (3000×35) are clustered using k-means with 200 clusters (200 clusters has been shown to be reliable for BoW representation in computer vision applications (Lazebnik et al., 2006; Scovanner et al., 2009)). After obtaining the cluster centres (codebook), each new cranial data is projected into the codebook to find the points that lie close to the cluster centres. By using max pooling the visual features for a cranial is summarized into 64 bins histogram. The results obtained using KELM, SVM, SRC classifiers are presented in Table 4.1. When the entire cranial is considered for classification with K-ELM classifier, the sex prediction accuracy attained by LD-SIFT is 50%, Spin Image attained 67.5%, and SIHKS provided 50% accuracy. This initial results show that the representation derived using long tailed feature vectors with bag of word model do not offer distinctive information to properly distinguish between the two class of sex.

Table 4.1: Sex prediction results using long tailed features with BoW representation

Feature Extraction	Classifier	Full (%)	3 Regions (%)	5 Regions (%)	7 Regions (%)
SIHKS	KELM	50	**67.5**	60	60
	SVM	51.1	**67.5**	60	55
Spin Image	KELM	67.5	60	56	62.5
	SVM	67.5	62.5	57.5	**68**
LD-SIFT	KELM	50	60	57.5	45
	SVM	50	**62.5**	57.5	53

4.3.2 Initial Analysis using Stacked Representation

Since BoW does not seem to be useful for cranial sex prediction, the second representation technique describe in section 4.2 is evaluated. For this purpose also, whole skull is initially used for feature representation (the prediction results are indicated in column "Full"), before partitioning the skulls in different sub-regions.

All experiments are evaluation using the 4 local feature descriptors. By examining the use of the full skull, but with the stacked representation of local descriptors from multiple points, LD-SIFT attained a prediction rate of 56%, Spin Image provided 69%, the result of SIHKS is 45.3%, while MeshLBP attained a prediction rate of 67.1%, as shown in Table 4.2. With SVM, the recognition generally improved by atleast 2.5% in most cases. For example, the result of Spin Image increased to 72.3%, while MeshLBP increased to 71.5%. This can be attributed to the minimization of correlation among features since all points are summarized into a single entity and also the ability of SVM to attain very good generalization on unseen test set.

Table 4.2: Sex prediction Results with stacked features

Feature Extraction	Classifier	Full (%)	3 Regions (%)	5 Regions (%)	7 Regions (%)
LD-SIFT	KELM	66	68.5	**76.5**	72.5
	SVM	63.5	68	**76.5**	73
	SRC	68	73	73.75	**74.5**
Spin Image	KELM	74	76	**80**	75
	SVM	74	74.5	**82**	76
	SRC	72.25	71.5	**80.5**	73
SIHKS	KELM	45.3	51.1	56.6	50
	SVM	46.2	48.9	54.4	56.7
	SRC	63.5	56.5	62	60.5
MeshLBP	KELM	63.5	74	**80.25**	70
	SVM	58	74.5	**82**	64
	SRC	62.5	73	**80**	60.5
LD-SIFT + KPCA	KELM	65	67.5	68	**72.5**
	SVM	68	75	**77.75**	74
	SRC	68	67.5	72	**74.5**
Spin Image + KPCA	KELM	77	73	**82.5**	75
	SVM	77.5	81.5	**83**	76.5
	SRC	71	73.5	**81.75**	73
SIHKS + KPCA	KELM	51.1	56	52.5	55
	SVM	53.5	47	52	53.5
	SRC	52	52	54.3	51
MeshLBP + KPCA	KELM	58.5	80.5	**85.25**	66
	SVM	59.5	77	**86**	70
	SRC	60.5	70	**81.3**	70
LD-SIFT + KSVD	KELM	67.5	74.5	75	67
	SVM	60	68	68.5	67
	SRC	55	64	68	64.5
Spin Image + KSVD	KELM	75	76	**83**	75
	SVM	60	74.5	**75.5**	71
	SRC	59.5	61	**76**	65
SIHKS + KSVD	KELM	44.5	57.6	51	60
	SVM	56	58	48	51.5
	SRC	61	55	49.5	59.5
MeshLBP + KSVD	KELM	49	69.5	**83**	70.5
	SVM	51	60.5	**80**	55
	SRC	51.5	65.5	**74.5**	55

To investigate the influence of different regions on prediction accuracy, the

multi-region representation approach described in Section 4.2 is explored. Using SRC and region partitioning into 3 regions, the prediction is between 56.5 and 73% for the 4 local descriptors. Using KPCA and KSVD for dimensionality reduction and feature learning the performance decreased by atleast 5% for all descriptors, as shown in Table 4.2. It was noted that probably no good subspace can be found to represent the data at lower dimension for 3 sub-region partitioning. However, by dividing the cranial into 5 sub-regions, significant improvement is observed in the performance of MeshLBP, LD-SIFT, and spin image, as shown in Table 4.2. In this case, the best results was attained using KPCA with spin image. In Fig. 4.5, an example of real testing scenario is illustrated, where the test sample is subject number 11 (male subject). The training matrix depicted on the right side of the figure demonstrate the sparse coefficients learned from the training matrix containing male and female subjects. Notably is that the largest coefficient value is found at the index of subject number 8 (also a male subject). This directly signifies that both samples belong to the same class. In this experiment, careful tuning of the regularization value resulted in better sparse vector but at the detriment of sex classification performance, hence the regularization value that provided the best performance was selected. However, despite the coefficients not being totally sparse, it still produced a classification performance of 81.75%. It can also be concluded from this experiment that 5 sub-region partitions is reliable enough for cranial sex prediction, as it can be seen in Table 4.2 that the results decreased drastically using 7 sub-regions.

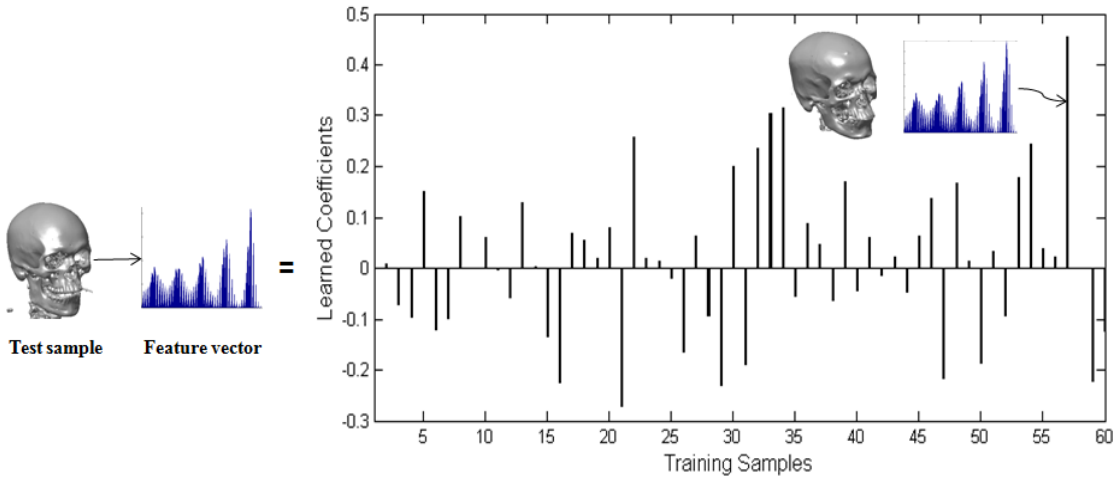
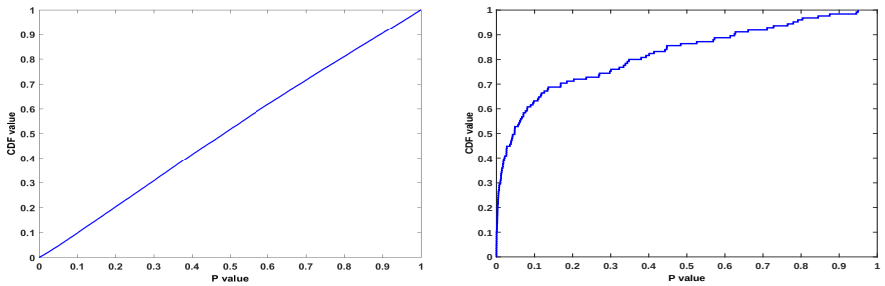


Figure 4.5: **A valid SRC test.** The skull on the left belongs to subject number 11, which is in the test set. The figure on the right illustrates the learned coefficients on the training matrix, with largest coefficient found as subject number 8. As a result, the test shows that these two samples belong to the same class (male).

Using KELM classifier on 3 sub-regions, the 4 descriptors provided a results between 51.1 and 76 %, with KPCA the results ranged between 56 and 80.5%, while KSVD yielded prediction rates between 57.6 and 74.5%. Similar to the case of SRC, the prediction results improves by dividing the skulls in 5 sub-regions with the best result 85.25% attained using MeshLBP and KPCA, as shown in Table 4.2. Moreover, it can be noticed that the performance decreased when the skulls are partitioned into 7 sub-regions. With SVM classifier, the results are quite close to that of KELM, and the best prediction 86% is attained using MeshLBP and KPCA on 5 sub-region of the skull, as highlighted in Table 4.2. In Section 4.3.3 to make the evaluation of the extracted 3D local descriptors comprehensive, the partitioning of the skull is performed along the X, Y, and Z axes, as depicted in Fig 4.4 and the number of regions, N examined in each axis is $N = \{1 : 2 : 99\}$.

In order to support the claim in Section 4.2 that stacked representations are more distinctive than long tailed representation, the two sample student's t-test evaluation is conducted on the extracted features from skull. The test is used to obtain a decision for the null hypothesis that the data in two feature vectors come from independent random samples from normal distribution with equal mean and unknown variance. The alternative hypothesis is that the two feature vectors come from distribution with unequal means. The p values from the test offer a sense of how the features can separate the two classes. Besides, to understand the visual interpretation of the result, the empirical cumulative distribution function of the p values can be plotted as depicted in Figure 4.6.



(a) (b)

Figure 4.6: Student's T-test conducted on extracted features (a) Long tailed feature representation (b) Stacked feature representation

The figures show that long tailed features are not distinctive enough for sex determination as the plot of cumulative distribution function against different p-values of t-test is linear, while the stacked feature representation indicates a higher CDF for small p-value of the t-test.

4.3.3 Effect of Number of Patches and Partition Orientation on Sex Prediction

This subsection demonstrates how the orientation of partition and the number of skull partition influence the results of sex prediction. These evaluations are divided into 2 categories, where the 1st evaluation involves using ordinary 3D local descriptors to train the 3 predictive models and the 2nd evaluations involves using KPCA and KSVD to learn compact representation from the feature vectors before classification. In the first evaluation, the skulls are initially partitioned into different regions, and the local features are extracted, aggregated and finally concatenated from all regions before serving as an input to the predictive models. It can be noticed in Fig. 4.7 that partitions derived from Z-axis generally provided the best prediction rates, particularly using MeshLBP as these results remain consistent across the 3 predictive models. When the MeshLBP features extracted from partitions in the Z-axis are used to train KELM classifier, the highest prediction rate obtained is 80.25 %, SRC classifier showed similar performance with a prediction rate of 80 %, while with SVM classifier the highest prediction rate is 82 %. Moreover, in all experiments, MeshLBP showed better performance than region area, ShapeDNA, SI-HKS, LD-SIFT, and Spin Image local descriptors.

Whereas, the Y-axis in some cases showed comparable trend in performances, especially using LD-SIFT and Spin image local descriptors, which also remain consistent across the 3 classifiers. For instance the prediction results of spin image are 76.25 %, 73.5 %, 73 % using KELM, SRC, and SVM classifiers respectively.

However, for MeshLBP it can be seen that the prediction results achieved from partitions in X and Y axes are lower that those from Z axis. Besides that, it is observed that the effect of increase in number of partitions is not so evident in all experiments, except for MeshLBP where the prediction results peaked at 5 regions before deteriorating as the number of partitions approach 99 regions. With region area, ShapeDNA and SI-HKS the performances are neither influenced by orientation of partition nor number of regions as the prediction rate is generally less than 70% in the 3 predictive models. Thus, leading to the conclusion that the SI-HKS descriptor is not suitable for sex prediction from human skulls.

In the second evaluation, KSVD and KPCA are used to learn compact representations from the local features prior to classification, which significantly reduce the dimensionality of the feature vectors. Using KSVD, a slight improvement in the performance of the local descriptors can be noticed. For instance, MeshLBP increased from 82% to 83% using KELM classifier, while the results using SRC and SVM also increase by atleast 1 %. Similarly, the prediction rates of LD-SIFT

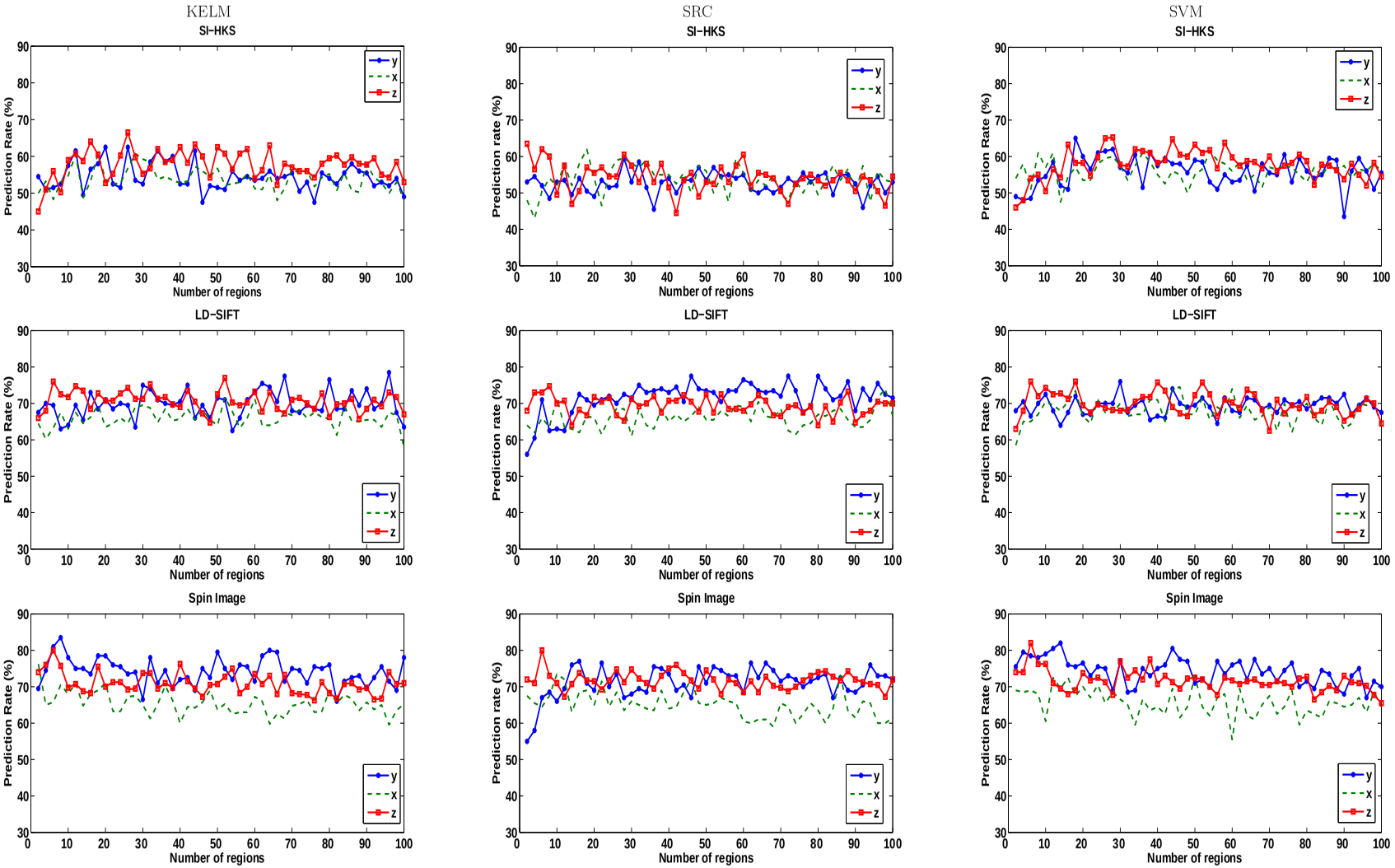
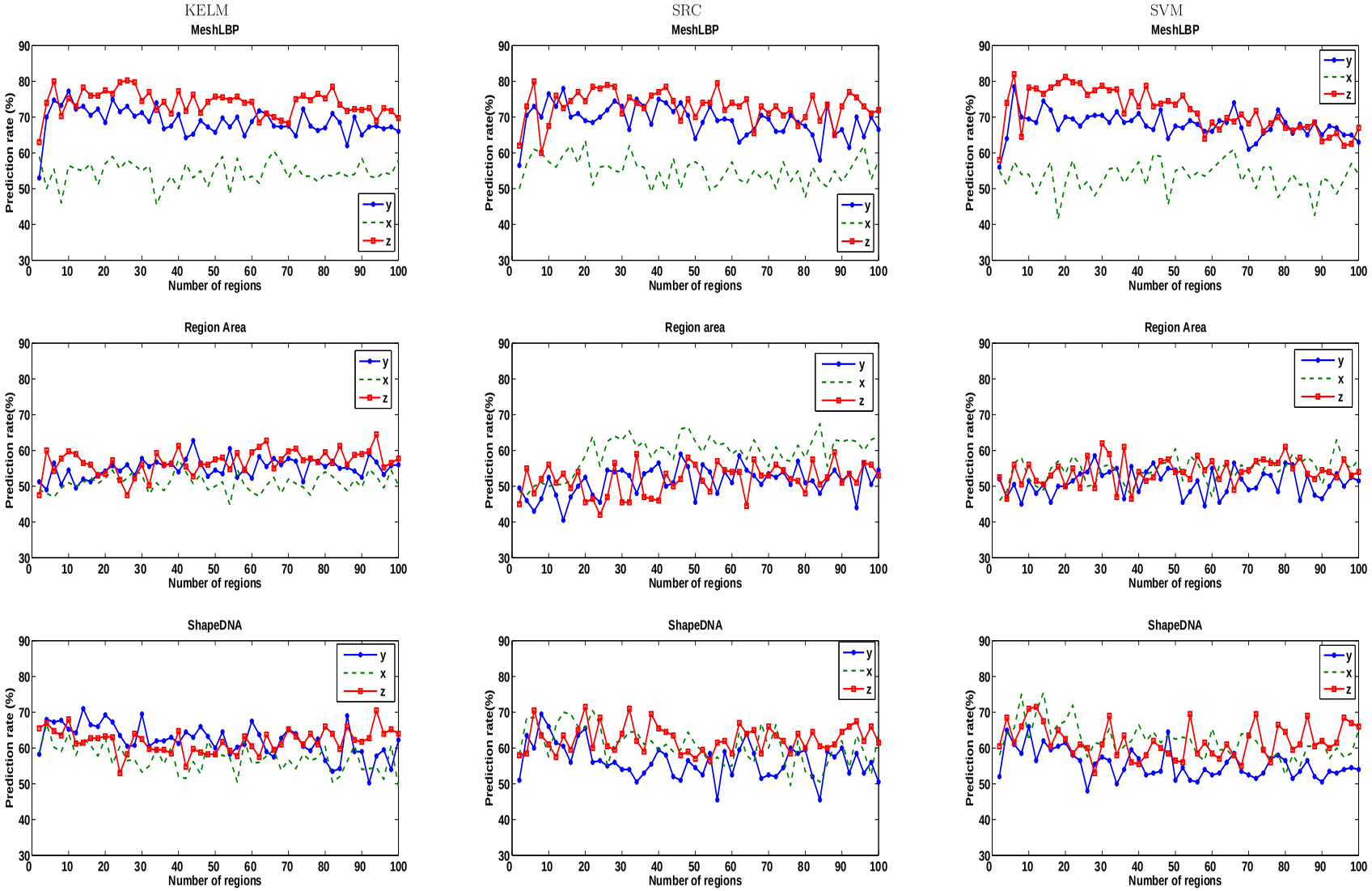


Figure 4.7: Results of sex estimation based on orientation of region partition and number of regions. Contd. on next page


KELM
SRC
SVM
MeshLBP
MeshLBP
MeshLBP
Region Area
Region area
Region Area
ShapeDNA
ShapeDNA
ShapeDNA
Prediction rate (%)
Prediction rate(%)
Number of regions
y
x
z

and Spin image increased across the 3 classifiers. Meanwhile, a similar trend in prediction results with respect to the orientation of partitions can be observed from the results in Fig 4.8, as the Y-axis tends to exhibit comparable performance to the Z-axis partitions.

Using KPCA, improvement in prediction results, especially in the case of spin image and MeshLBP where the prediction rates increased to 85.5% and 86% respectively can be observed, as depicted in Fig. 4.9. In addition, it is obvious that the partitions derived from the Z-axis provided better results than X and Y-axes. Also, increasing the number of partitions resulted in reduced prediction results for MeshLBP, however the trend is only slightly evident for spin image and LD-SIFT. Overall, SVM classifier showed better prediction performance than SRC and KELM.

4.3.4 Sex determination using MKL on Multiple Features

The essence of the experiments in this section is to investigate the possibility of using MKL to integrate information from multiple features. As such, in this experiment, the linear kernel function is used for mapping each local feature, which are then combined in the MKL algorithm. The focus is placed on LD-SIFT, Spin image, and MeshLBP since these 3 feature representation methods showed better performance than region area, ShapeDNA and SI-HKS in Section 4.3.3. Using this algorithm 4 different sets of combination are performed with respect to the orientation of partition (X,Y,Z) and number of regions ($1 : 2 : 99$). Particularly, the following fusion have been examined:

- LD-SIFT + MeshLBP
- Spin Image + MeshLBP
- Spin Image + LD-SIFT
- Spin Image + LD-SIFT + MeshLBP

The results attained are illustrated in Fig 4.10. Learning a multi-kernel representation for LD-SIFT + MeshLBP the best prediction rates attained are 72.5 % (X-axis), 71.5 % (Y-axis), and 78 % (Z-axis). With Spin image + MeshLBP the prediction rates are 72 % (X-axis), 70 % (Y-axis), and 86 % (Z-axis), while the results of Spin Image + LD-SIFT are 79.5 % (X-axis), 72.5 % (Y-axis), and 85.5 % (Z-axis). Similar to the case of single descriptor, it can be noted that the partitions derived from the Z-axis provided better performance than X and Y axes in all experiments.

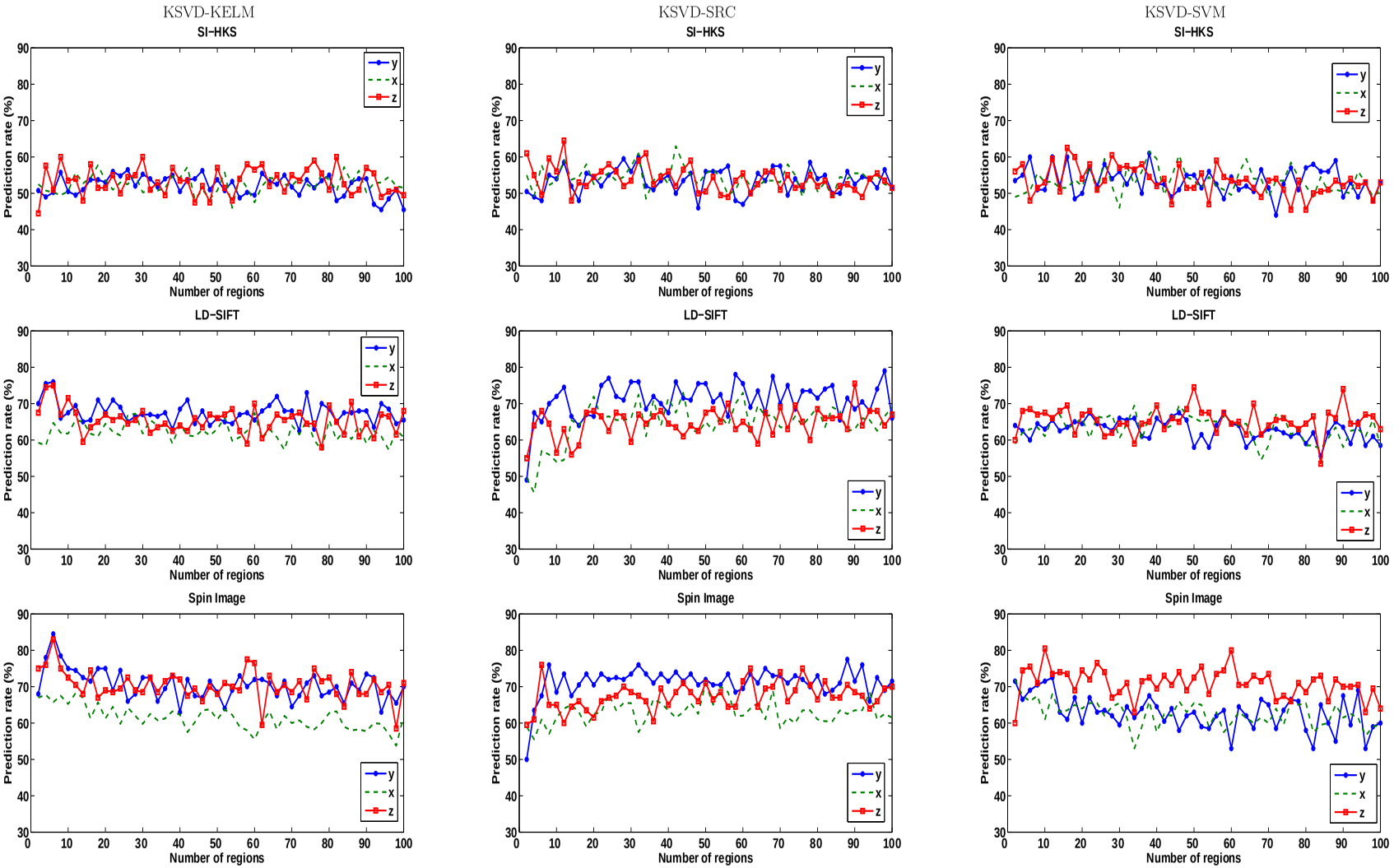
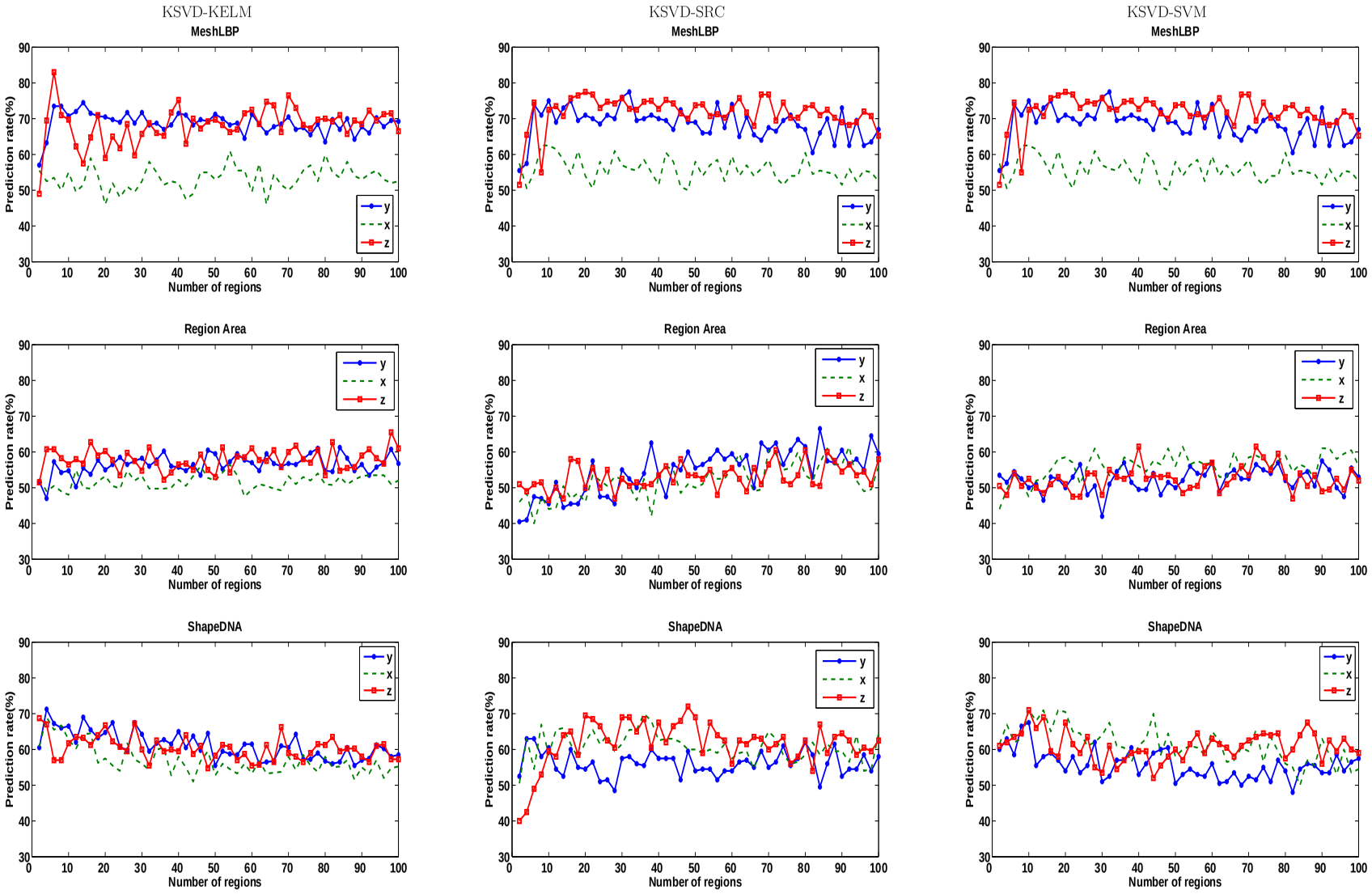


Figure 4.8: Results of sex estimation based on orientation of region partition and number of regions using KSVD. Contd. on next page


KSVD-KELM
KSVD-SRC
KSVD-SVM
MeshLBP
Region Area
ShapeDNA
Prediction rate(%)
Number of regions
y
x
z

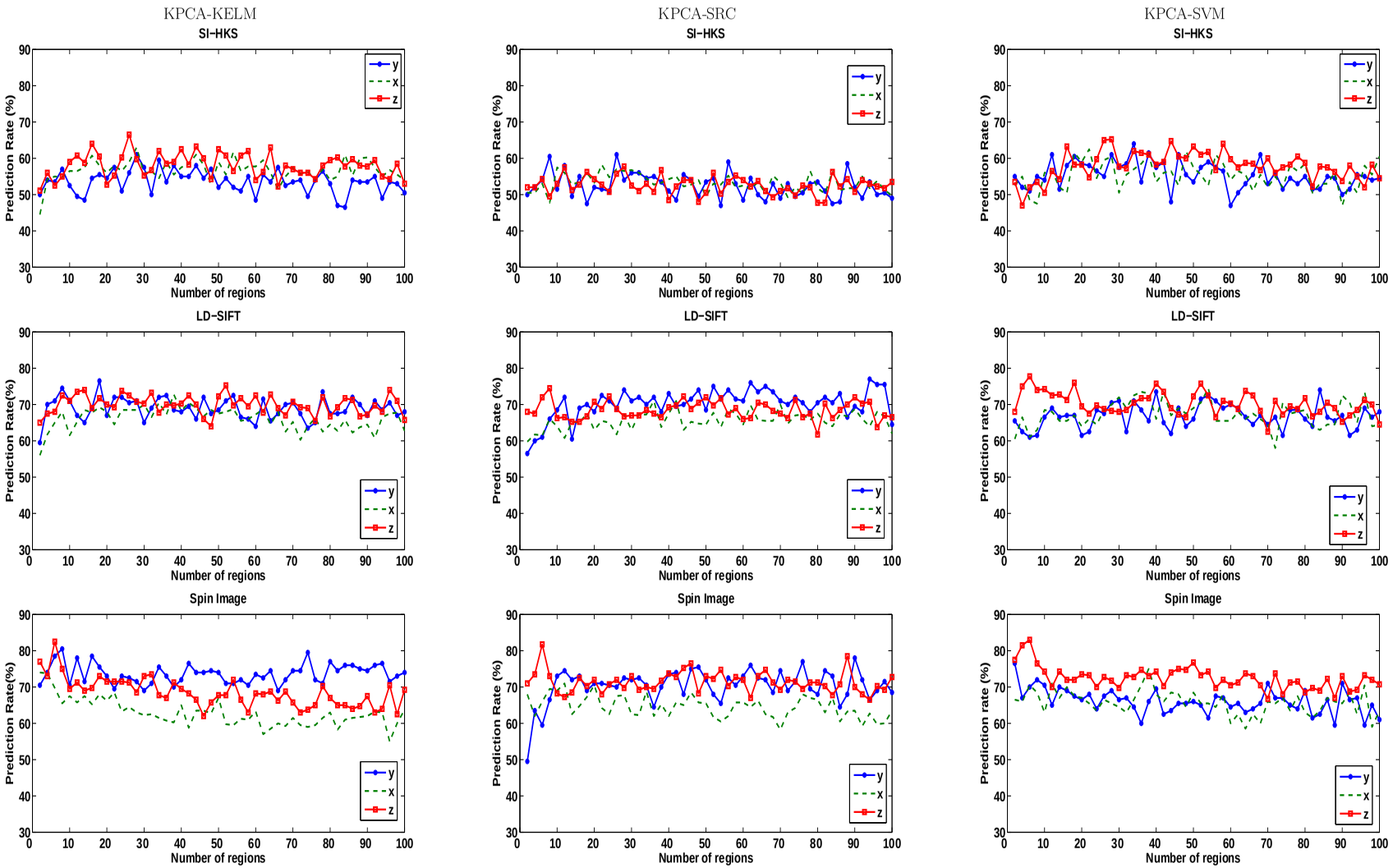
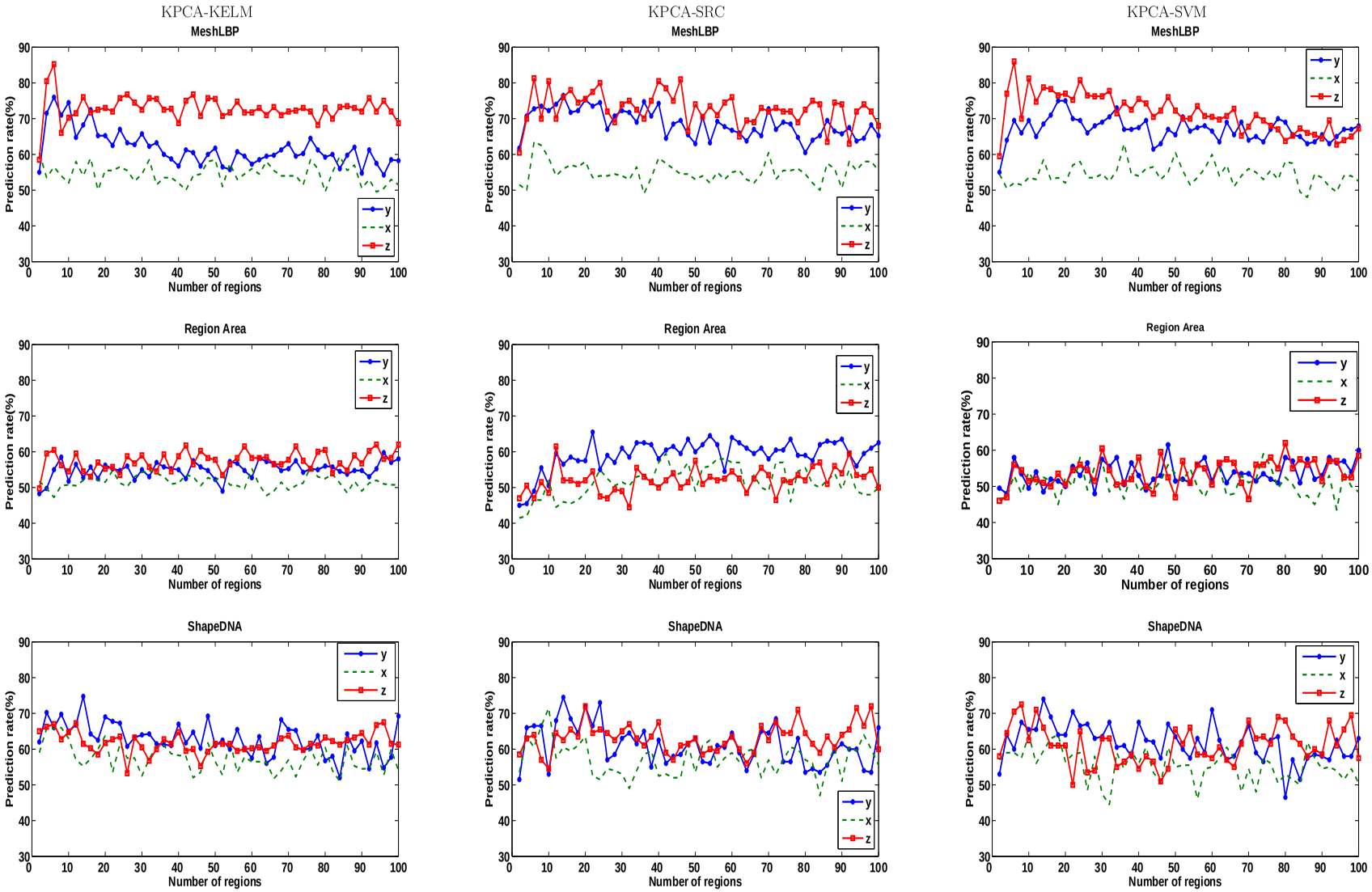
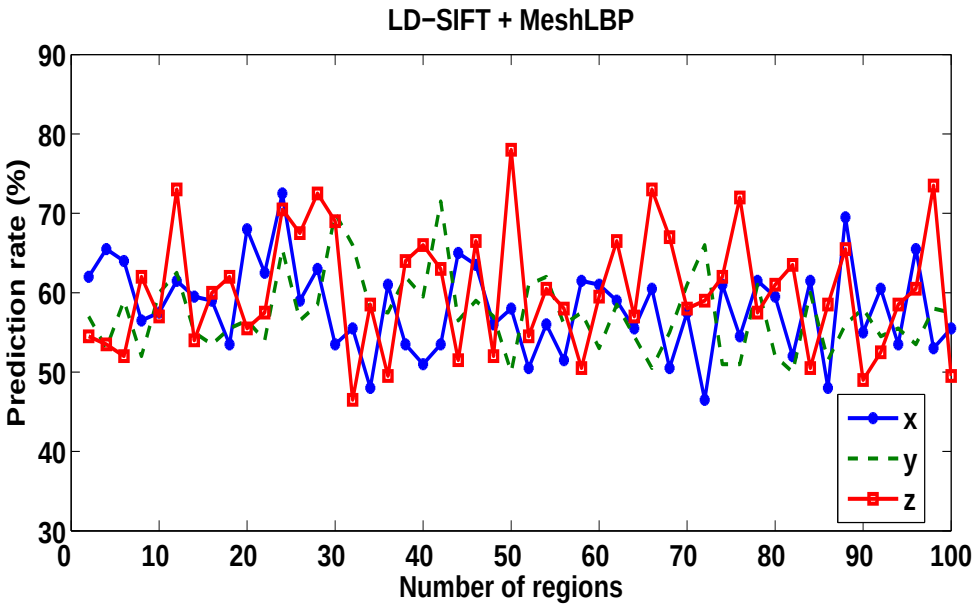
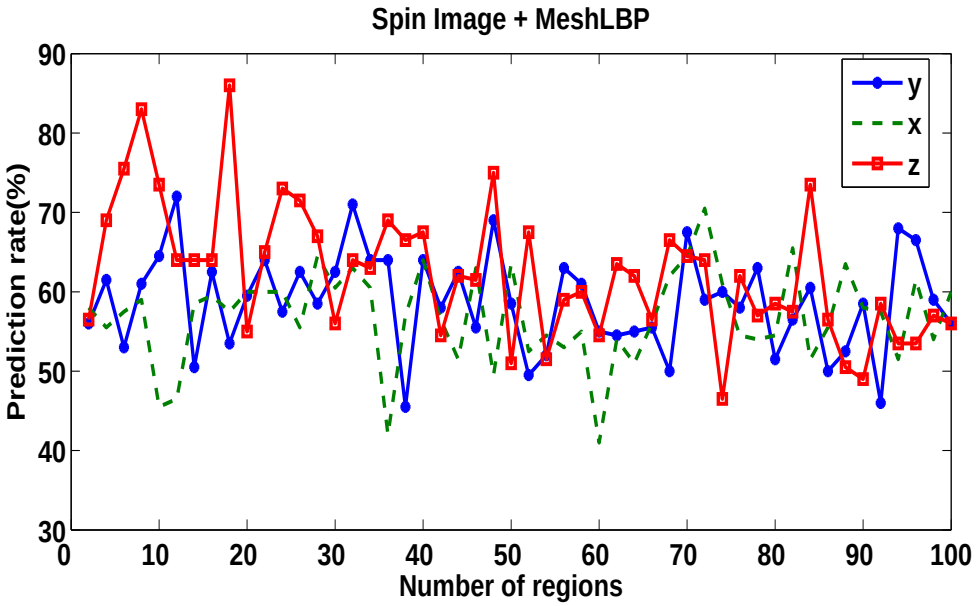


Figure 4.9: Results of sex estimation based on orientation of region partition and number of regions using KPCA. Contd. on next page


KPCA-KELM
KPCA-SRC
KPCA-SVM
MeshLBP
Region Area
ShapeDNA
Prediction rate(%)
Number of regions
y
x
z


LD−SIFT + MeshLBP
Prediction rate (%)
Number of regions
x
y
z


Spin Image + MeshLBP
Prediction rate(%)
Number of regions
y
x
z

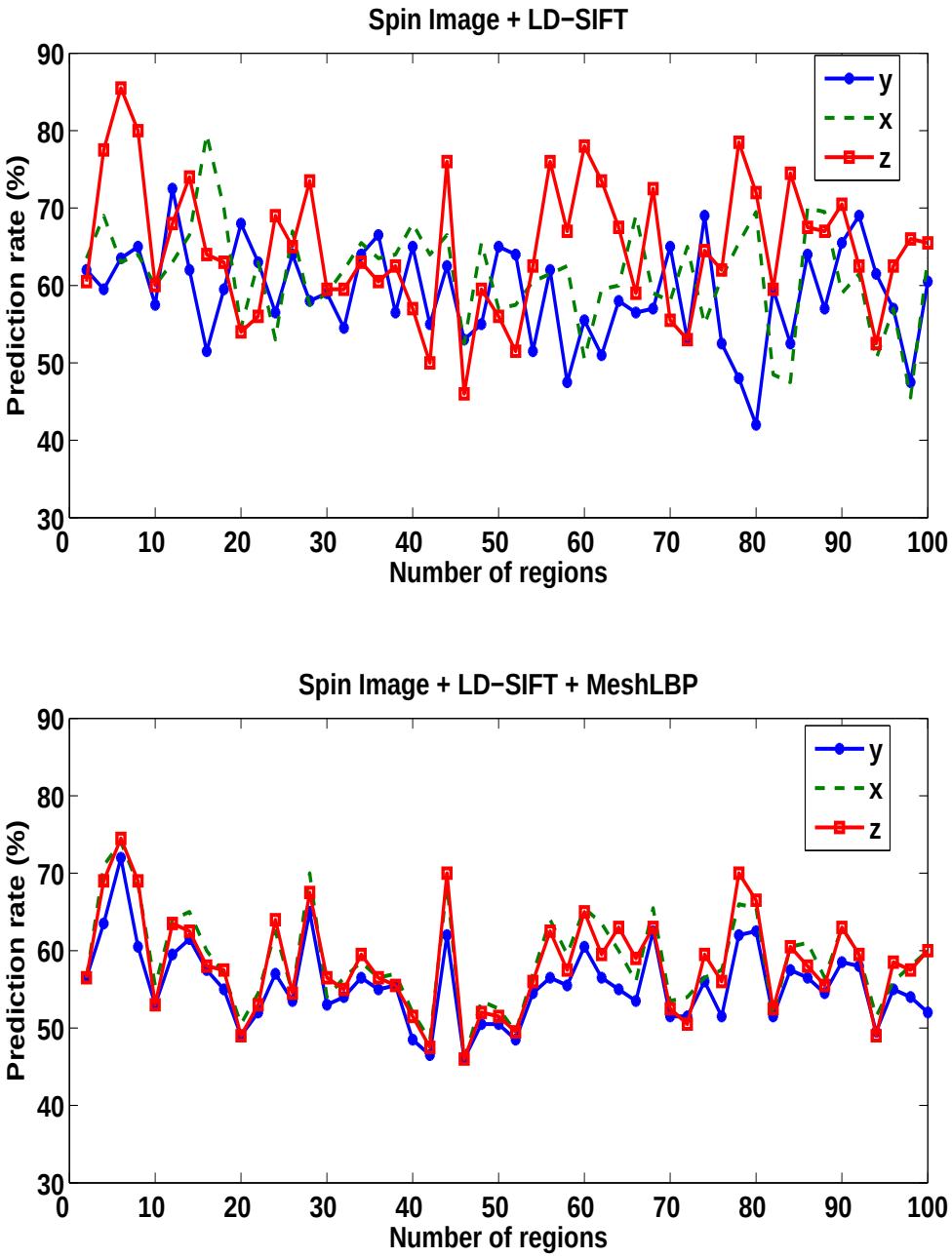


Figure 4.10: Results of sex estimation based on orientation of region partition and number of regions using MKL

Moreover, the influence of number of regions is slightly evident on the results of Spin image + LD-SIFT and Spin + MeshLBP, as the peak recognition rate is achieved at 5 regions before receding as the partitions approach 99 regions. However, no significant impact of number of region can be observed on the prediction rates of LD-SIFT + MeshLBP. Finally, the combination of Spin image + LD-SIFT + MeshLBP is attempted, but the prediction results only decreased on the 3 orientation of skull partition, indicating less compatibility among the 3 features.

Quite interestingly, Spin + MeshLBP produced the most compatible combination as the prediction rate of 86 % is comparable to the benchmark attained with single descriptor. It can be denoted from these experiments that MKL is useful for sex prediction from human skulls albeit the selection of the most compatible 3D local descriptor seems to be necessary. Unlike the soft margin MKL which tries to induce information from multiple sources, the hard margin MKL based on primal formulation, which imposes penalty on the features and selects those that best optimize the objective is also tested. However the results as shown in Figure 4.11 are not so promising compared to the soft margin MKL. This can be attributed to the fact that the hard margin MKL tends to be dependent on the discriminative ability of the base kernels (Gehler and Nowozin, 2009). Thus, in a case where the single features are already discriminant, the hard margin MKL will be unable to exhibit any better performance than the single features. On the other hand, this indicates that if the single features are not discriminant such as heat kernel signature or ShapeDNA, their combination will have a negative impact on the performance unless they can be singled out.

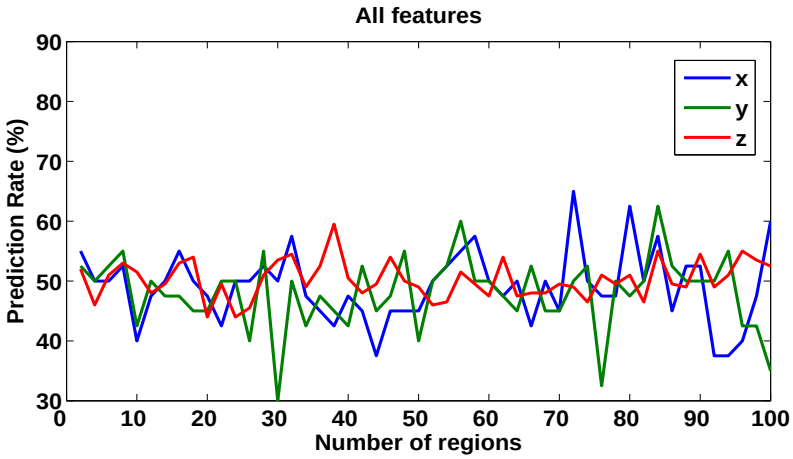


Figure 4.11: An example of sex identification using hard margin MKL with primal formulation for 3D feature combination.

4.4 Conclusion

In this chapter, several analyses of sex determination techniques based on patch based representations has been presented. A technique for deriving patch representation of mesh points from different skull orientation and at different multi-levels, covering large patches (3 patches) to very small patches (99 patches) has been introduced and evaluated using feature learning methods and predictive models. The results indicate that patch based features derived from partitioning the skull are

suitable for sex identification from 3D skulls, with the representation derived using MeshLBP producing the best prediction rate of 86%. Partitioning the skull along the Z-axis tends to produce the best prediction results (since it is less affected by imperfect skull alignment), although other orientations (X and Y) provided very competitive results. Moreover, it is observed from the results that the effect of number of partitioning (patches) on the prediction results has no statistical significance. Furthermore, empirical evaluation of combination or fusion of different feature representations using MKL attest to the possibility of learning information from more than one source. Similarly, it is observed that number of partition, as well as orientation of partition have no significant effect on the prediction results.

The empirical evaluation in this Chapter is essential for gaining profound understanding of the dimension in which sex determination can be approach using CV based methods. Classic forensic based methods make use of manually localised and quantified information, which inherently makes challenging issues such as improper alignment/registration, unwanted noise, and comprehensive feature representation negligible. However, it is counter-intuitive because the developed system will have to operate manually in stages with immense human intervention. To function automatically will require addressing those challenging issues. For instance, noise removal is necessary in an automated system to eliminate the influence of outliers. Furthermore, this Chapter has shown through empirical evaluation the usefulness and weakness of different local features, as well as the effect of the representation strategy (whether long tailed representation or stacked representation), which has not been considered previously in the literature. Also, it has demonstrated through several experimental evaluations that accurate skull alignment/registration is not necessary, as it can be seen that the results from varying axes are competitive.

Chapter 5

A Pseudo Anatomical Region based Patch Representation

> You can't solve a problem on the same level that it was created. You have to rise above it to the next level
>
> Albert Einstein

This chapter presents the further experiments conducted on sex identification from 3D skulls by representing local patches of the skull according to the skull anatomical structure. The proposed approach in this chapter involves 3D skull preprocessing, skull partitioning into pseudo-anatomical regions, 3D feature representation and classification. Due to the fact that 3D preprocessing and feature extraction have been discussed in Chapter 3, thus the focus in this chapter is directed to how the skull can be partitioned into local regions which share similarity to anatomical structures.

5.1 Overview

With understanding from Chapter 4 that different orientations of skull partition produced prediction results that are rather similar in trend, it will be worthwhile to introduce and investigate a different approach of partition which is more geometrically sound, covering all 3 orientations at a time, as well as, incorporating the anatomical structure of the skull. The aim of this Chapter is to:

- Establish correspondence among skull samples through 3D registration.
- Examine the performance of sex determination using structural skull partitioning which is referred to as pseudo-anatomical partition.

The proposed approach initially commence with 3D preprocessing, where the input volumetric PMCT samples are filtered to remove noise and isosurface (mesh) reconstruction to obtain the hard tissue (skull layer). Skull partitioning is performed by first registering the skull samples to a model skull using coherent point drift (CPD) to establish correspondence and using fuzzy c-means to cluster the skull into different regions. Then 3D local shape features are extracted on the partitioned regions to derive a set of compact and discriminant representations. Finally sex determination is performed using different predictive models as shown in Figure 5.1.

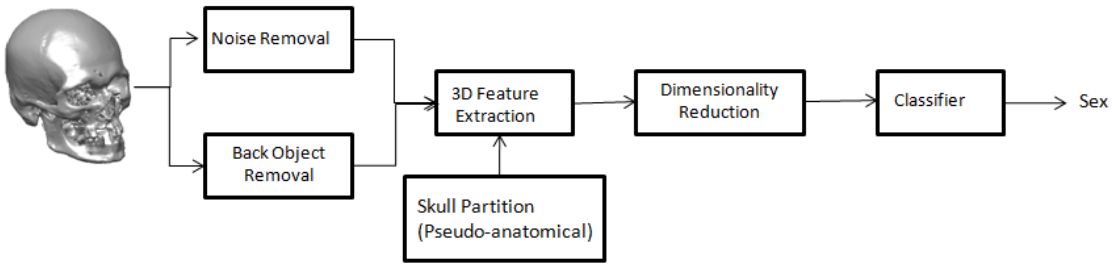


Figure 5.1: Structure of sex determination using anatomically partitioned skulls

5.2 Related Works on Skull Partitioning

In understanding the structural details of different skull regions, it is essential to deploy computational methods for partitioning the 3D skull into small regions. This is unlike the traditional or conventional forensic approach where manual mode of segmentation or partitioning is usually adopted prior to such analysis. Manual segmentation is the common practice particularly in morphological assessment (Walker, 2008). However, from the literature there are limited number of research reporting an automated mode of skull partitioning (AbdelFatah et al., 2013; Musilov et al, 2017). Nevertheless, there are a number of instance where 3D mesh requires partitioning into several sub-regions. As such, 3D mesh partitioning is rather well covered in object recognition than forensic analysis due to the necessity in that research domain (Attene et al., 2006). Due to the large body of literature on mesh segmentation, a number of notable classic approaches are discussed in this Section.

Region regrowing is one the commonly used concepts in mesh segmentation. Algorithms in this realm normally start by selecting a few seed points, which can be chosen randomly or computationally determined using some geometrical criteria. Each region is then segmented by expanding the corresponding initial seed

(Attene et al., 2006). Vieira and Shimada (2005) proposed a region growing algorithm which initially uses the noise and curvature of each vertex to partition the mesh into regions with fundamentally different characteristic shapes and then use these regions to create seed regions for region growing. The aim is to maximize the number of connected vertex points approximated by a single underlying surface at each vertex point by alternating between region growing and surface fitting. The process is finally completed by filling segment holes created by outlier noise. The curvedness of 3D mesh, which is invariant to rotation and translation has been computed as a geometrical criteria (Jagannathan and Miller., 2007). To obtain the segmented parts, the curvedness values are used to adaptively determine the curvedness thresholds for segmentation by iterative morphological operation composing of graph dilation and morphological filtering of outlier vertex points. Shapira et al. (2008) proposed the used of shape diameter function (SDF) as a criteria which captures the link between an object's volume and its boundary. The motivation behind this idea is to create a low pass filtering to the shape-diameter measure by examining the diameter in the neighborhood of each point on the surface. Vo et al. (2015) proposed a coarse to fine point cloud segmentation approach. The authors initially converted the point clouds into voxel representation and the voxels are grouped into adjacent voxels possessing similar characteristics. Then, region growing is performed on the octree-based voxelized representation of the input point cloud to extract the coarse segments and the output is passed through a refinement process. In a different approach, segmentation can be achieved by hierarchical processing of object's sub-part. Lai et al. (2008) extended the random walks algorithm to 3D mesh segmentation. The approach depends on 2-stage process, where the first stage first uses a set of automatically selected seed points to oversegments the mesh, and then merges these initial regions to produce the final regions. While region growing seem to be easy to implement, the main limitations of region growing are the dependency on multiple criteria and selection of seed points, which has no generally acceptable form of definition.

An alternative mode of segmentation is through clustering and one of the most commonly used algorithms in this domain are fuzzy clustering, k-means, and GMM (Katz and Tal., 2003; Sharma, 2009; Tsuchie, 2014). Clustering based techniques for segmentation are robust since they do not require seed points and the effect of the chosen geometrical criteria is highly minimal as the clustering itself can be performed directly on the mesh points.

Some authors have focused on using spectral analysis for mesh segmentation, which still remains an active area of research. Spectral analysis basically involves

the analysis of the eigenvalues and eigenvectors of a defined matrix. segmentation based on spectral analysis can be considered as the use of clustering on a graph which composes of vertices and edges in a mesh. From the on set the algorithm requires construction of an affinity matrix which will be used to derive a graph Laplacian matrix, which subsequently leads to solving the eigendecomposition problem on the laplacian matrix. Different authors have shown diverse methods of obtaining the laplacian matrix. For instance Liu and Zhang (2004) performed spectral clustering on 3D meshes by solving for the eigenvectors of a distance-based affinity matrix and segmented the mesh into k submeshes using clustering such as k-means on the largest eigenvalues of the normalized affinity matrix. The approach was further modified, where a partial construction of affinity matrix is required (Zhang and Liu, 2005). The algorithm is integration of recursive spectral two-way cut and Nystrom approximation. McGraw (2015) proposed a simple and flexible technique to mesh segmentation based on non-negative matrix factorization (NMF). Similar to the case of spectral analysis, the authors constructed an affinity matrix to obtain the geometric properties of the mesh. Spectral based segmentation is intuitively sound from algorithmic point of view. However, the main challenge is the issue of computational complexity, especially when attempting to solve the eigendecomposition for a large 3D mesh.

5.3 Pseudo-anatomical Partitioning

Pseudo anatomical skull partitioning is an automated partitioning of the skull into sub-regions that *partially* represent the local *anatomical* sites of the skull. *Partially anatomical* means the regions are not exactly the anatomical sites in standard forensics per se (since they are not annotated by forensic experts), but they reflect the geometric locations of those regions on a 3D skull. This inherently enables the possibility of approaching the partitioning as a standard clustering problem, which further gives the liberty to express those partitions from coarse to fine scale.

5.3.1 Coherent Point Drift

In order to achieve the goal of pseudo-anatomical partitioning, skull registration is first performed using coherent point drift (CPD) (Myronenko and Song, 2010) based algorithm. CPD uses a probabilistic method to align the centroids of gaussian mixture models (GMM) of two independent datasets, such that when a sample optimally aligns to the other, the maximum of their posterior probabilities is reached. Let $\mathcal{M}_{|\mathcal{M}|\times d} = \left(u_1, u_2, \ldots u_{|\mathcal{M}|}\right)^T$ be the model 3D points and

$\mathcal{T}_{|\mathcal{T}|\times d} = \left(u_1, u_2, \ldots u_{|\mathcal{T}|}\right)$ the target 3D points, where d is the dimension of the 3D points and $|\mathcal{M}|$ and $|\mathcal{T}|$ indicate the number of vertices in the two 3D data respectively. The operation of CPD is to compute the weighted GMM probability density function by taking extra outliers into consideration as (Myronenko and Song, 2010):

$$p(\mathcal{T}) = \omega \, \frac{1}{|\mathcal{T}|} + (1-\omega) \sum_{|m|=1}^{|\mathcal{M}|} \frac{1}{|\mathcal{M}|} p(\mathcal{T} \mid \mathcal{M}) \tag{5.1}$$

where $p(\mathcal{T} \mid \mathcal{M}) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{||\mathcal{T}-\mathcal{M}||^2}{2\sigma^2}\right)$. Hence, the algorithm uses the expectation-maximization (EM) method to find the 3D transformation parameters $\Theta = (\phi, \sigma)$.

During the **E**-step of i-th iteration of the EM algorithm, the posterior probability of GMM is computed using (Min et al., 2016):

$$p^{(i)}(\mathcal{M} \mid \mathcal{T}_{|\mathcal{T}|}) = \frac{\exp\left[-\frac{1}{2} \parallel \frac{\mathcal{T}_{|\mathcal{T}|} - Tr(\mathcal{M}_{|\mathcal{M}|}, \phi^{(i)})}{\sigma^{(i)}} \parallel^2\right]}{\sum_{|m|=1}^{|\mathcal{M}|} \exp\left[-\frac{1}{2} \parallel \frac{\mathcal{T}_{|\mathcal{T}|} - Tr(\mathcal{M}_{|\mathcal{M}|}, \phi^{(i)})}{\sigma^{(i)}} \parallel^2\right] + (2\pi(\sigma^{(i)})^2)^{\frac{d}{2}} \frac{\omega}{1-\omega} \frac{|\mathcal{M}|}{|\mathcal{T}|}} \tag{5.2}$$

where $Tr(\mathcal{M}_{|\mathcal{M}|}, \phi^{(i)})$ is the transformation parameters of the GMM centroids. In the **M**-step, a new set of parameters is derived by solving for the upper bound limit of a log-likelihood function $L(\Theta)$ (Min et al., 2016; Myronenko and Song, 2010):

$$L(\Theta) = log\left[\prod_{|t|=1}^{|\mathcal{T}|} p(\mathcal{M}_{|\mathcal{M}|})\right] = \sum_{|t|=1}^{|\mathcal{T}|} \log \sum_{|m|=1}^{|\mathcal{M}|+1} p(\mathcal{M}) \; p(\mathcal{T}_{|\mathcal{T}|} \mid \mathcal{M}) \tag{5.3}$$

Computationally, the complexity of CPD is $\mathcal{O}(|\mathcal{M}|^3)$ (Myronenko and Song, 2010), which means that the computation increase dramatically as the number of 3D vertices increase. Considering the 3D skulls used in this thesis, the number vertices even after downsampling (in Section 3.2.1) is close to 130,000 vertices, thus CPD is not directly useable for the 3D skull data.

5.3.2 Ensemble CPD

The original CPD is not implementable directly on a huge point set such as the 3D skull considered in this research. As such, a new computational improvement is proposed to the CPD using random data sampling without replacement (taking into consideration the spatial contiguity of the data points) and batch-by-batch registration.

Sampling without replacement involves selecting a sample at random from a population and the sample is not returned to the main lot (in this case, the vertices of the 3D skull is referred to as the population and a vertex point is a sample). The first sample is selected out of the population of size $|\mathcal{M}|$ and the second sample is selected out of the remaining population of $|\mathcal{M}| - 1$ samples, which is continued until all samples are selected. In other words, there is a decrease in the population size as the sample size $|m|$ increases. More importantly, the sample size $|m|$ cannot be greater than the population size $|\mathcal{M}|$ and once a sample has been selected, it can not be chosen again. Hence, all the selected samples are different to one another.

To improve the CPD, the original vertex points of skull are divided into different batches using random sampling without replacement. The random sampling is performed, taking the spatial contiguity of the points into consideration. Then, the algorithm uses a global tolerance function which minimizes till it converges to a particular global error value or reaches the maximum number of iterations. In order to register a particular, the CPD algorithm is used to register the two data samples as $[err, transf, corr] = CPD(\text{sample1}, \text{sample2})$, where err is the iteration error of registration, $transf$ is the transformation matrix, and $corr$ is the correspondence index. The iteration error err is derived as a function of the distance between the model and registered sample, which is then divided by the maximum distance and tunable tolerance parameter τ that controls the rate of convergence of the error. This process is performed for all the batches until the maximum number of iterations is reached or the error is greater than the preset tolerance values. The procedure of the proposed improved CPD is as illustrated in **Algorithm 1**:

Algorithm 1 Ensemble CPD

Input: Model sample $\mathcal{M}$, Target sample $\mathcal{T}$, maxitr, tolerance, τ
Divide $\mathcal{M}$ into $\{\mathcal{M}_1, \mathcal{M}_2, \ldots \mathcal{M}_J\}$ subsets and $\mathcal{T}$ into $\{\mathcal{M}_1, \mathcal{M}_2, \ldots \mathcal{M}_J\}$ batches
for each itr **do**
  for all J, $\mathcal{R}_J \leftarrow \mathcal{T}_J$ **do**
    for each subset j **do**
      $[err, transf, corr] = CPD(\mathcal{M}_j, \mathcal{R}_j)$
      $\mathcal{R}_j = transf * \mathcal{R}_j$
      $D = dist(\mathcal{M}_j, \mathcal{R}_j)$
    end for
    $err = \frac{\sum D}{Max(D)*\tau}$
  end for
  if err $\geq \tau$ **then**
    Stop iteration
  else if $itr \equiv$ max itr **then**
    Stop iteration
  end if
  All corr $= [corr_1, corr_2, \ldots corr_J]$
end for
Output: All corr

In order to demonstrate the effectiveness of the proposed method, 5 sample 3D shapes are selected from the Shape Analysis Research Project (SHARP 3D) [1] dataset, which are (Bunny, 3D face, Cartilage, Vase, and Rocker Arm) to compare Ensemble CPD with standard CPD and non-rigid ICP. The results in terms of registration time and error are illustrated in Figure 5.2 and Tables 5.1 and the convergence error with respect to the iterations in Figure 5.3.

[1] https://www.itl.nist.gov/iad/vug/sharp/

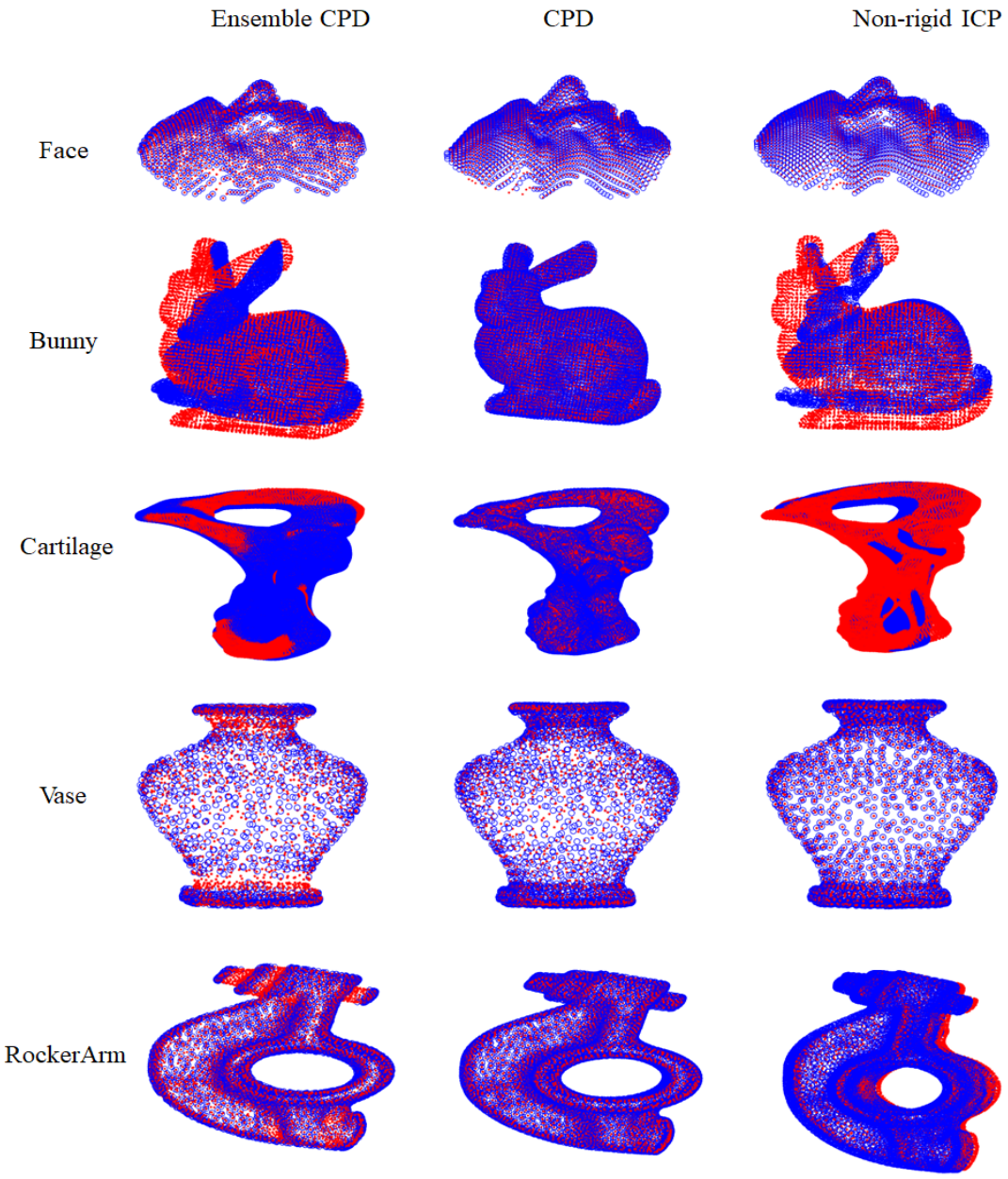


Figure 5.2: Comparison of registration methods. Blue color indicates the source (model) data, while red color is the registered data.

Table 5.1: Computational time of registration

Methods	Face	Bunny	Cartilage	Vase	Rockerarm
Non-rigid ICP	0.86	0.84	15.87	0.17	1.73
CPD	3.85	23.09	34.07	2.36	12.21
Ensemble CPD	0.18	0.89	2.54	0.44	1.7

Table 5.2: Registration error based on distance between model and registered target

Methods	Face	Bunny	Cartilage	Vase	Rockerarm
Non-rigid ICP	0.0163	0.0923	1.618	0.017	0.014
CPD	0.0150	0.070	1.607	0.0523	0.065
Ensemble CPD	0.0159	0.091	1.743	0.0268	0.083

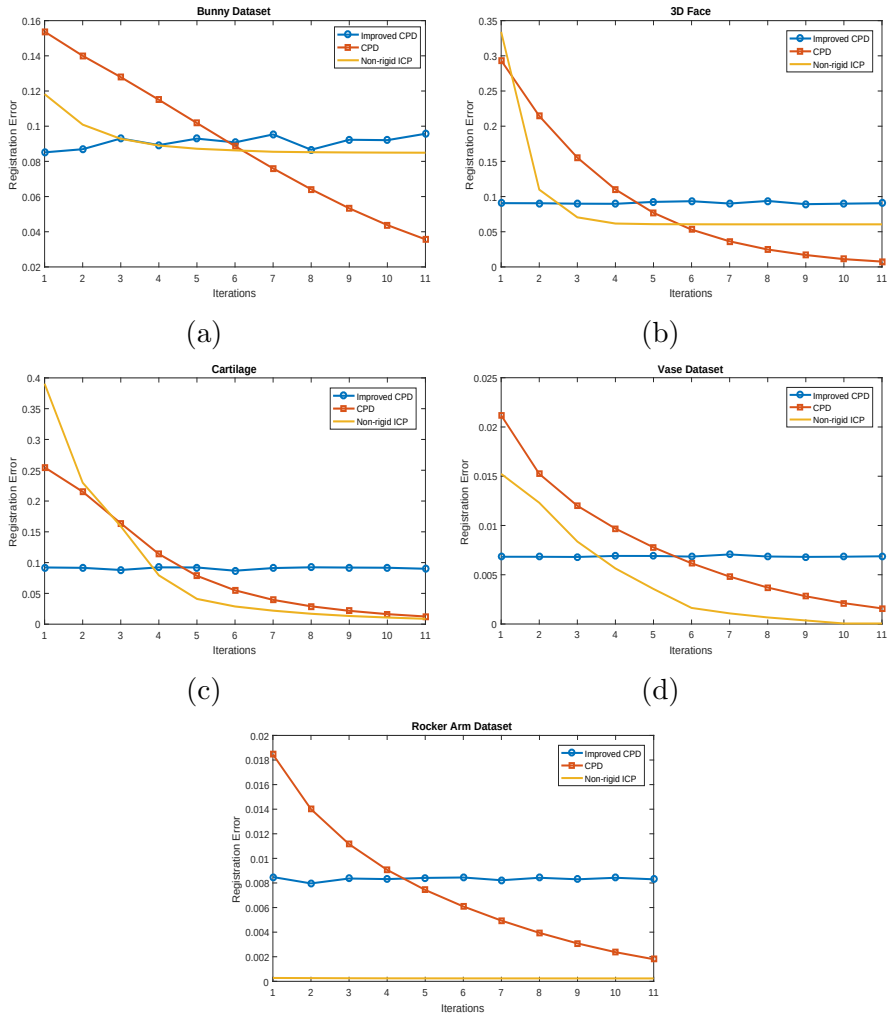


Figure 5.3: Registration error with respect to iterations on the 5 datasets.

The proposed skull registration method outperforms non-rigid ICP and improves the original CPD in terms of computational cost and time as shown in Table 5.1. In fact the original CPD and non-rigid ICP are not usable on huge point set as the computational requirement significantly exceeds the memory capacity of the PC.

All cases of experiments on large point set returned an out of memory error [2].

[2]The computer used for the experiments in this thesis is a workstation which runs on Quad

However, using the data splitting approach of the proposed method such challenges have been mitigated and the registration time is greatly improved. However, such improvement is at the detriment of the registration error, as it can be noticed that the registration error of the proposed method is higher than CPD as shown in Table 5.2. This is because the solution provided by the proposed algorithm is sub-optimal since the original data has already been divided into smaller subset reducing the original intrinsic structure of the data.

5.3.3 Application of Ensemble CPD to Skull Registration and Partitioning

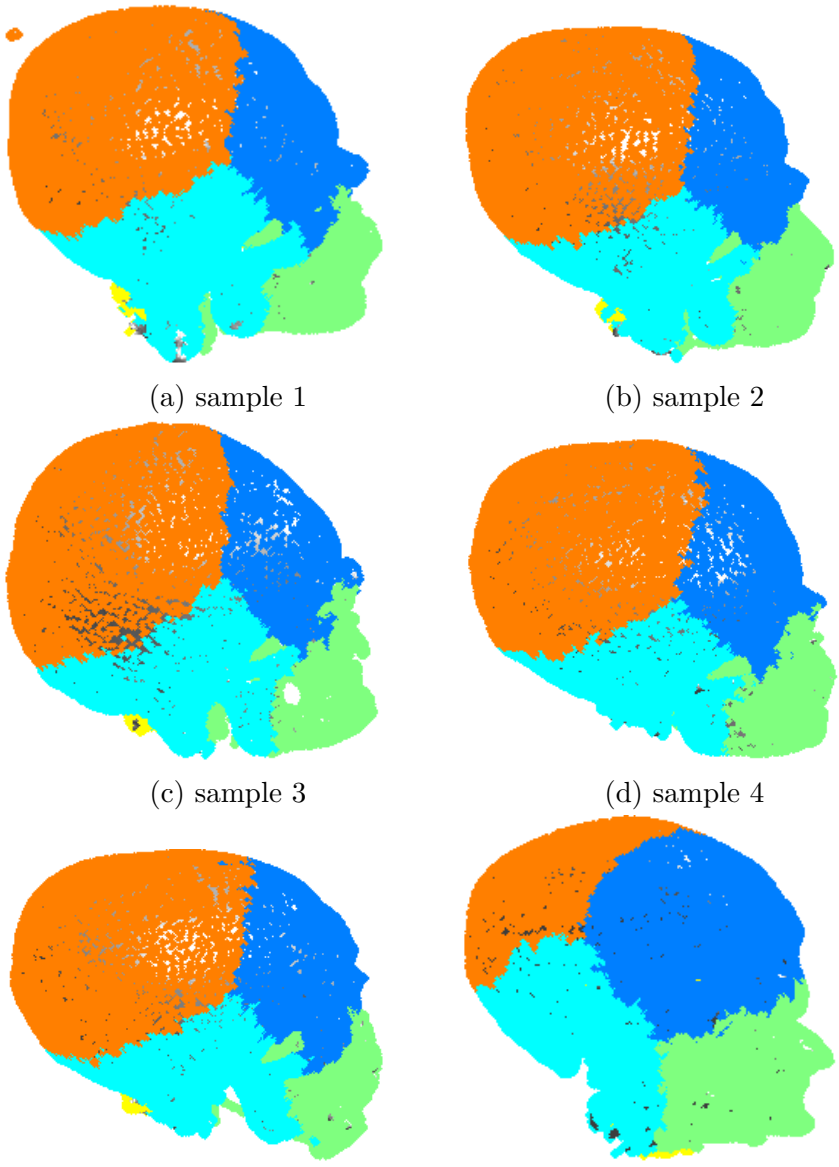
During registration, a model skull is used to register each of the remaining sample set to obtain their correspondence. The skull registration is first performed for the skull samples using the improved coherent point drift (CPD) Myronenko and Song (2010). Let $\mathcal{M} = \left\{u_1, u_2, \dots u_{|\mathcal{M}|}\right\}$ and $\mathcal{T} = \left\{u_1, u_2, \dots u_{|\mathcal{T}|}\right\}$ be the model skull point set and target skull point set respectively, where $u_i, i = 1, \dots |\mathcal{M}|$ and $v_i, i = 1, \dots |\mathcal{T}|$ are 3-dimensional coordinates of points, and $|\cdot|$ represents the cardinality of a set. Without loss of generality, it is assumed that $|\mathcal{M}| \leq |\mathcal{T}|$. $C : \{1, \dots, |\mathcal{M}|\} \rightarrow \{1, \dots, |\mathcal{T}|\}$ can be defined as one-to-one mapping such that $\forall k \in 1 \dots, |\mathcal{M}|, \quad \exists \ell \in 1 \dots, |\mathcal{T}|, \ F(U_{C(k)}) = V_\ell$, or simply denote $C(k) = \ell$, where F denotes a registration function such as CPD. Note that, if $C(k_1) = \ell_1$, $C(k_2) = \ell_2$, and $k_1 \neq k_2$, then $\ell_1 \neq \ell_2$.

Afterward, the model skull is selected for pseudo-anatomical region partitioning by clustering the mesh points using Fuzzy c-means (FCM). Once the model skull has been partitioned, the remaining samples are similarly partitioned by propagating the clusters from the model skull to the remaining samples. Assume that $\mathcal{M}$ is partitioned into N clusters, i.e $\mathcal{M} = \bigcup_{i=1}^{N} \mathcal{M}_i$, then the target is partitioned at least into N sets $\mathcal{T}_i \subseteq \mathcal{T}, \quad i = 1, \dots, N$ such that $\mathcal{T}_i = \{V_\ell \in \mathcal{T} \ \mid \forall k, \ U_k \in \mathcal{M}_i, \ C(k) = \ell\}$. Examples of pseudo-anatomical skull partitions on different skulls are shown figure 5.4.

5.3.4 Feature Extraction on Skull Partitions

From each region (partition) the 3D local shape features such as MeshLBP, Spin Image, and LD-SIFT are computed and aggregated. For instance, to extract MeshLBP in a local region, each mesh point takes a set of ring neighbors (10 rings are used in this thesis). Then, the local difference between the primitive function (e.g gaussian curvature) of a mesh point and that of its ring neighbors is

Core Intel Xeon with a memory of 8GB RAM.



(a) sample 1 (b) sample 2

(c) sample 3 (d) sample 4

(e) sample 5 (f) sample 6

Figure 5.4: Pseudo-anatomical Skull partitions of 6 different samples showing the locations of 5 regions (patches) of the skulls.

calculated as expressed in Eq 3.14. The MeshLBP is computed for r number of rings. Computing this for all mesh points $\mathcal{T}$ result in a $|\mathcal{T}| \times 10$ binary pattern as in **Algorithm 2**.

Algorithm 2 Pseudo-anatomical partition and feature extraction

Input: Model skull $\mathcal{M}$, Target skull $\mathcal{T}$, Number of scales N
Cluster $\mathcal{M}$ into N clusters using FCM, obtaining $\{M_1, M_2, \ldots M_N\}$
for $i = 1, \ldots N$ **do**
  find $\mathcal{T}_i$ in $\mathcal{T}$ according to $\mathcal{M}_i$
end for
for $\mathcal{T}_i, i = 1, \ldots N$ **do**
  for $j = 1, 2, \ldots \dim(\mathcal{T}_i)$ **do**
    Compute MeshLBP using (3.14) for P_j *in* $\mathcal{T}_i$
  end for Aggregate to form $MeshLBP(\mathcal{T}_i)$
end for
$RegionLBP = [MeshLBP(\mathcal{T}_1), MeshLBP(\mathcal{T}_2), \ldots MeshLBP(\mathcal{T}_N)]$
Output: $RegionLBP$

The patterns serve as the feature descriptors. In each sub-region $\mathcal{T}_N$ the ring descriptors are stacked on each other and the binary features are aggregated along each ring into 32-bin histogram, making a $32 \times 10 = 320$ local features [3]. Finally, the resulting feature descriptors from each partition (sub-region) are concatenated (making $320 \times N$ feature vector) as depicted in Figure 5.5 for classifying the skulls into different sex groups. Furthermore, kernel principal component analysis (KPCA) and k-singular value decomposition (K-SVD) are adopted to reduce the dimension of the features and predictive models such as SVM, SRC, and KELM are used for classification.

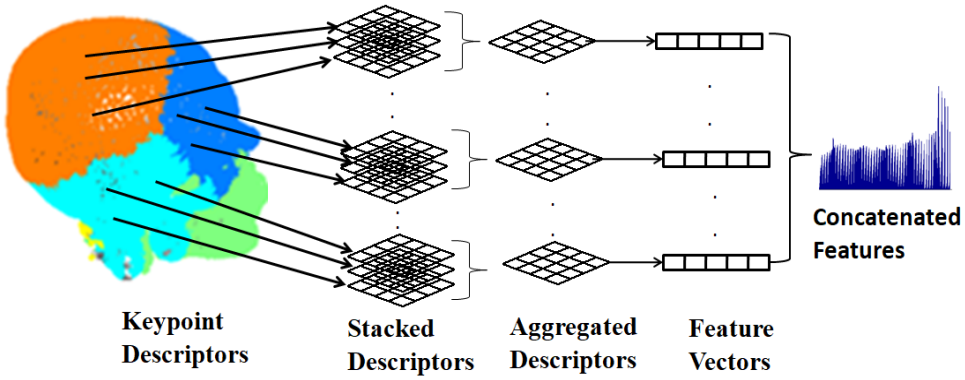


Figure 5.5: Illustration of patch representation of 3D local shape descriptors from pseudo-anatomical regions.

[3]The aggregation is based on stacking the features on one another and computation the summation

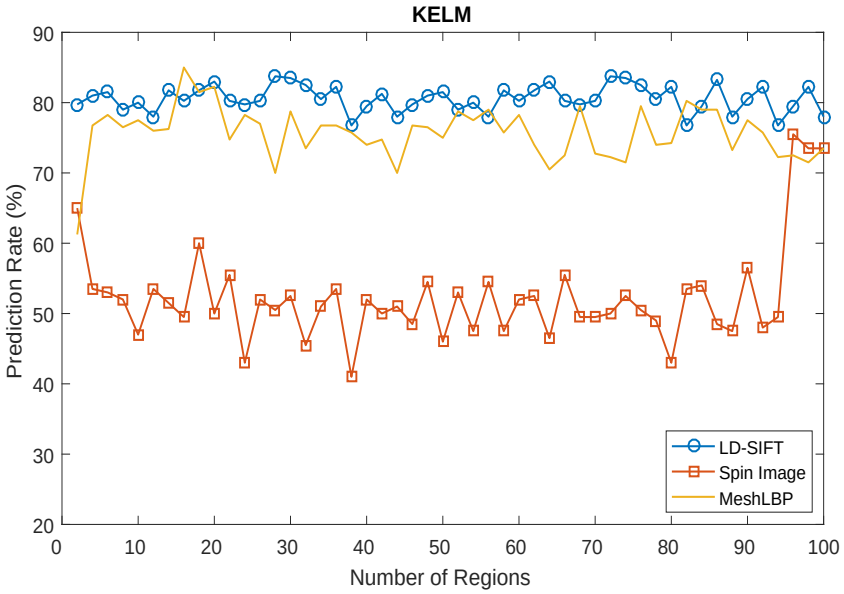
5.4 Sex Determination from Pseudo-anatomical Patches

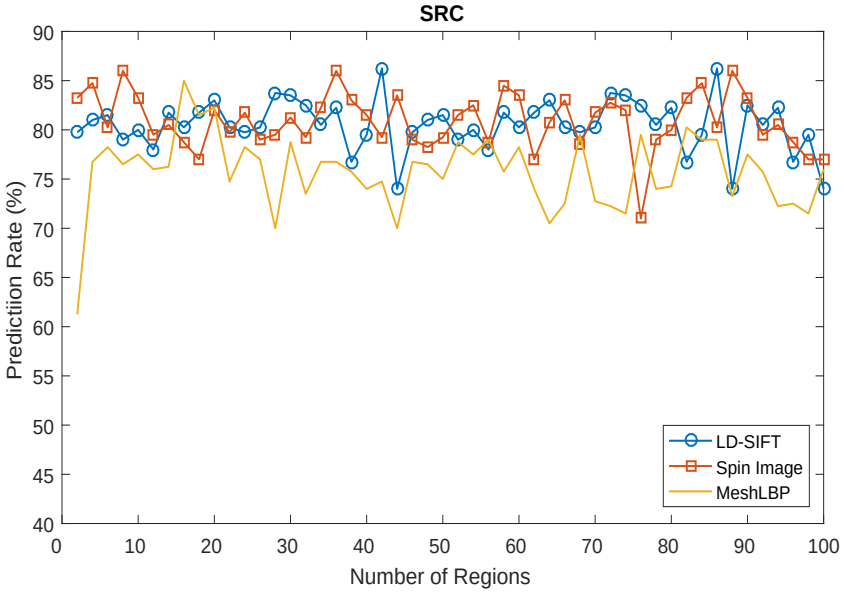
This section presents the experimental results and performance evaluation of using 3D local shape descriptors for forming patch based representation from the derived pseudo-anatomical regions, and classifying the sex of human skull. The experiments are conducted on 100 samples, where 60% of the dataset are utilized for training and 40% for testing. During the training phase a 5-fold cross validation is performed to determine the best regularization parameter, where $\lambda = [2^{-10} : 2^{10}]$. After discovering the best regularization value, the model is tested with the remaining 40% testing samples. All experiments are repeated 10 times and the average is computed as the final sex prediction result.

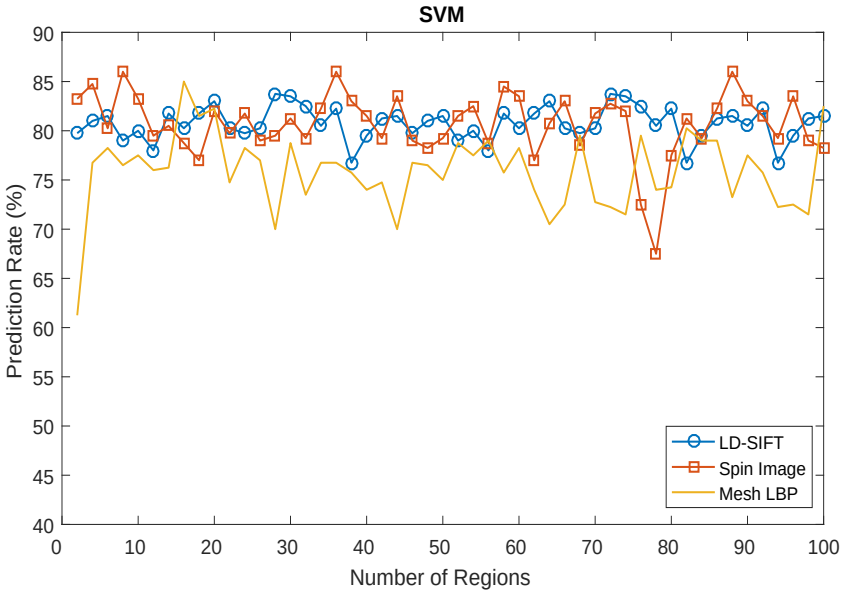
5.4.1 Effect of Number of Pseudo-anatomical Regions on Sex Determination

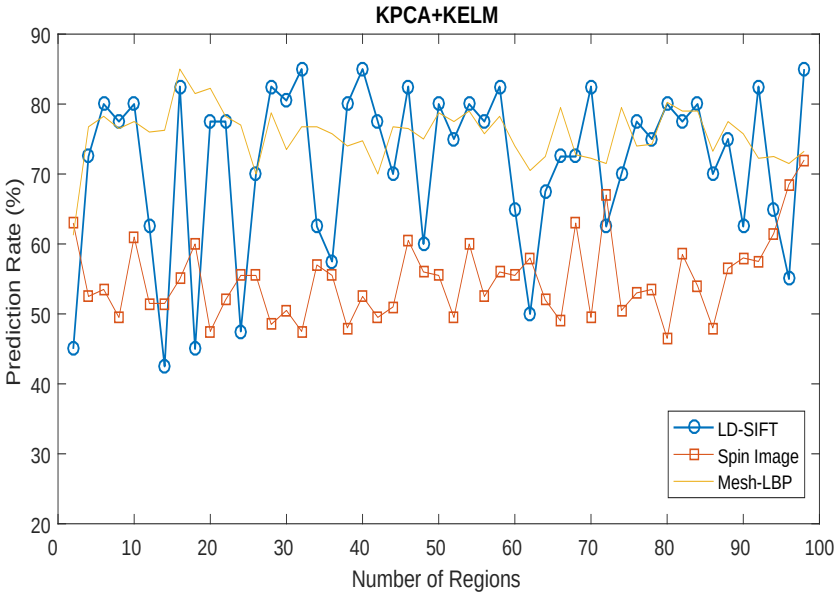
This section of thesis demonstrates the influence of number of pseudo-anatomical regions the skull is partitioned on sex determination, as well as the effect of the choice of 3D local descriptors (MeshLBP, LD-SIFT, and Spin Image) and classification techniques. In light of this, each skull sample is partitioned into N regions from $\{1 : 2 : 99\}$. Thus, each skull in the dataset is partitioned into different regions, and a 3D local descriptor is computed for each vertex point in the region, which are then aggregated and concatenated to form the feature vectors as shown in Figure 5.5

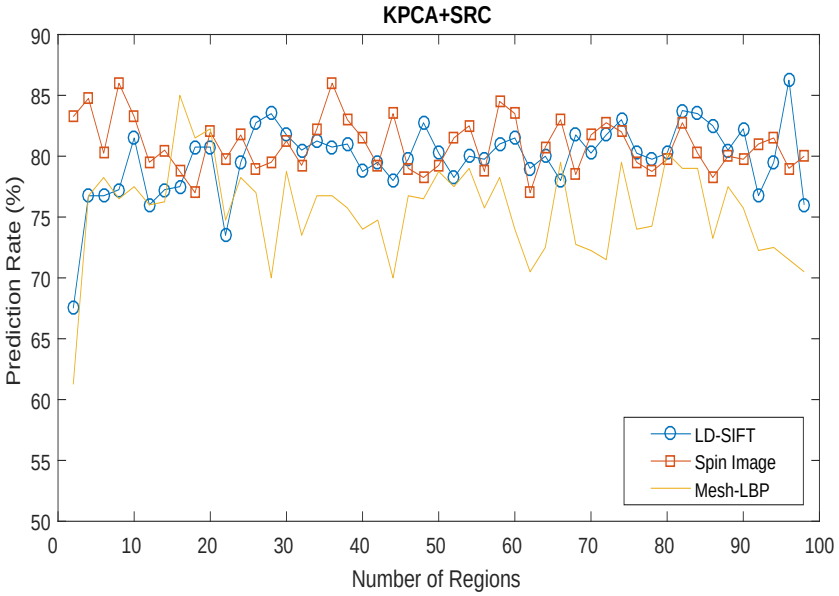
Initially, the original 3D local shape descriptors are used to train and test the 3 predictive models (SVM, KELM, and SRC) and the results are depicted in Figure 5.6. From the experiments, it is observed that LD-SIFT produced the most impressive performance with sex determination result of 85.7% using SVM classifier, while the best results of MeshLBP and Spin Image are 85% and 75.5% obtained using SVM and KELM classifiers respectively. Subsequently, the 3D local shape descriptors are processed using dimensionality reduction algorithms such as KSVD and KPCA to improve the compactness and representative nature of the features. The experimental results show that using MeshLBP with KPCA and SVM classifier, a sex determination performance of 86% can attained, Spin Image with KPCA and SVM also produced a prediction rate of 86% and LD-SIFT provided the best prediction performance of 86.5% with KPCA and SVM. The results attained with SRC and KELM classifiers using either KPCA and KSVD are generally not as good as SVM classifier.

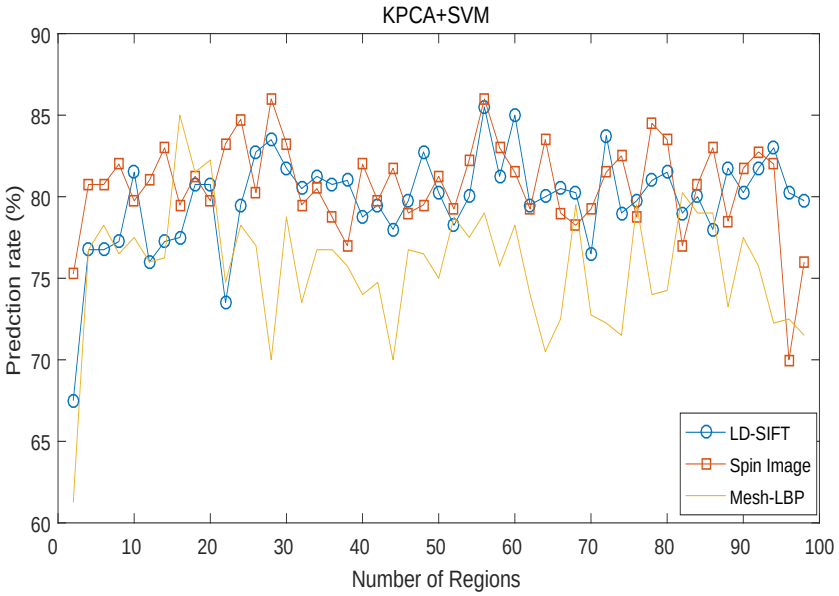

KELM
Prediction Rate (%)
Number of Regions
LD-SIFT
Spin Image
MeshLBP

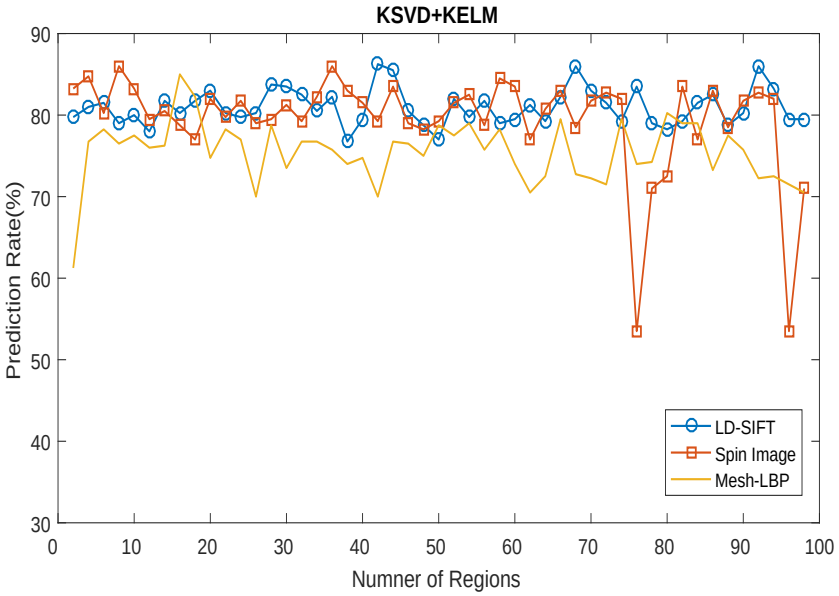

SRC
Predictiion Rate (%)
Number of Regions
LD-SIFT
Spin Image
MeshLBP

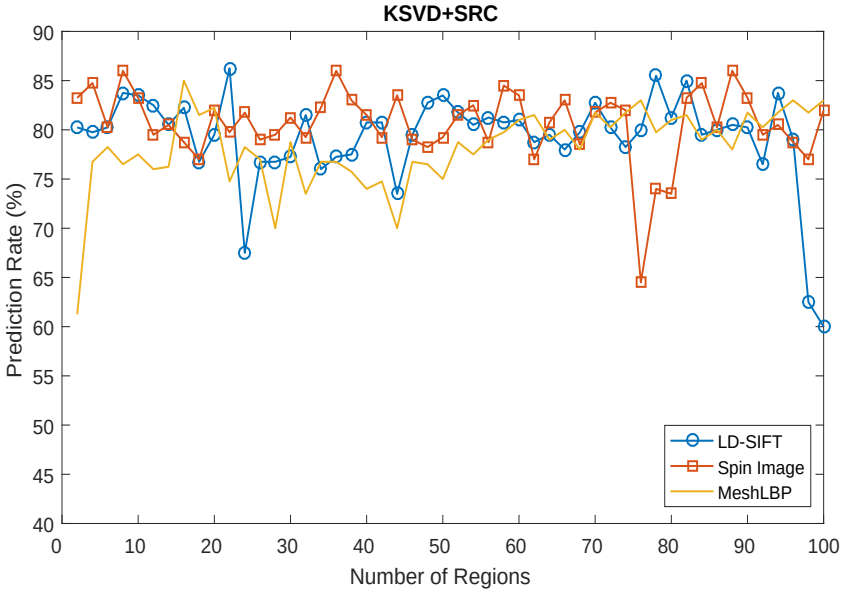

SVM
Prediction Rate (%)
Number of Regions
LD-SIFT
Spin Image
Mesh LBP


KPCA+KELM
Prediction Rate (%)
Number of Regions
LD-SIFT
Spin Image
Mesh-LBP


KPCA+SRC
Prediction Rate (%)
Number of Regions
LD-SIFT
Spin Image
Mesh-LBP


KPCA+SVM
Predction rate (%)
Number of Regions
LD-SIFT
Spin Image
Mesh-LBP


KSVD+KELM
Prediction Rate(%)
Numner of Regions
LD-SIFT
Spin Image
Mesh-LBP


KSVD+SRC
Prediction Rate (%)
Number of Regions
LD-SIFT
Spin Image
MeshLBP

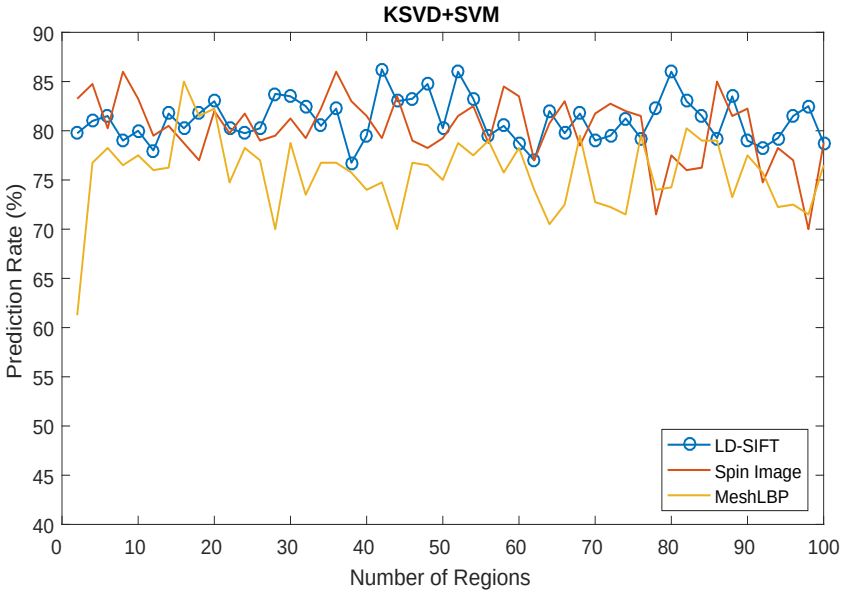


Figure 5.6: Sex prediction results for different number of regions (1 - 99) from pseudo-anatomical partitions using different feature extraction and classification methods.

In summary, the prediction results obtained in these experiments exhibit performance trend which are similar to the one in Chapter 4. The number of regions considered for partitioning the skull does not have significant impact on the prediction results. Also the feature representation methods appear to produce competitive results, though spin image and LD-SIFT slightly outperformed MeshLBP. This can be attributed to the fact that spin image and LD-SIFT have better strategy of encoding orientation information.

5.4.2 Sex Determination using MKL

This section presents the experiments conducted using MKL algorithm to integrate multiple features for sex determination. The theory behind MKL has been presented in Section 3.5.4. Thus, this experiment focuses on fusing the 3 local shape descriptors: LD-SIFT, Spin image, and MeshLBP, where a linear kernel function is used for mapping each local feature, which are then combined in the MKL algorithm. In essence, 4 different forms of combination as listed below are performed to derive the sex prediction results.

- LD-SIFT + MeshLBP
- Spin Image + MeshLBP
- Spin Image + LD-SIFT

- Spin Image + LD-SIFT + MeshLBP

In accordance to the mode of evaluation performed in Chapter 4 and Section 5.4.1, the number of regions the skull is partitioned into is considered as the basis of evaluation in this section, while also examining the prediction outcome of each feature combination. The results attained in the experiments are illustrated in Figure 5.7. Through the combination of LD-SIFT + MeshLBP using MKL framework the best prediction rate attained is 82.5% , Spin image + MeshLBP provided a result of 83, Spin image + LD-SIFT provided the best prediction result with an accuracy of 85%, while the prediction result of Spin Image + LD-SIFT are 82 %. This shows that the combination of spin image and LD-SIFT is more compatible than using MeshLBP. This can also be attributed to the share similarity in the encoding strategy of the two local features, for instance, both features are keypoint based and have a mechanical for computing keypoints' orientation and magnitude information in the feature descriptor.

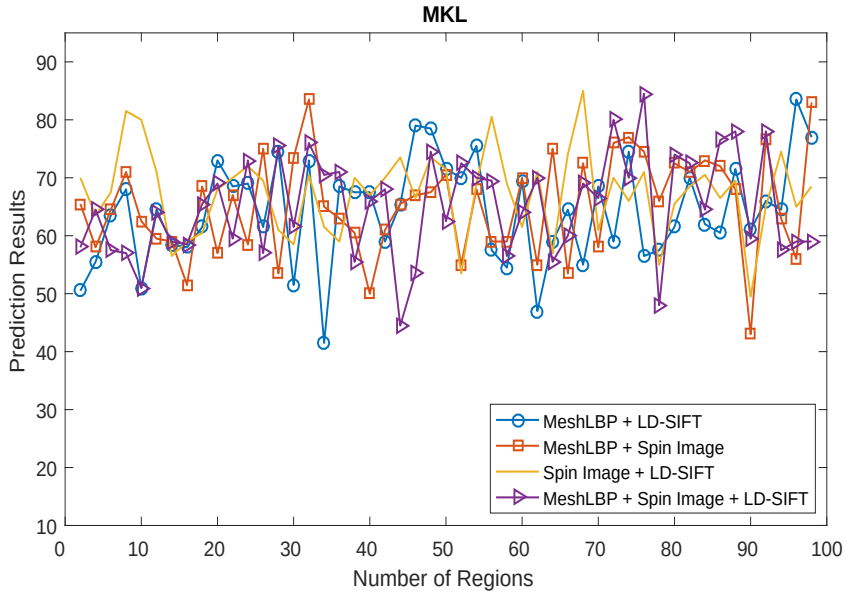


Figure 5.7: Illustration of patch representation of 3D local shape descriptors from pseudo-anatomical regions.

5.5 Conclusion

This chapter has presented a new method of skull partitioning named pseudo-anatomical partitioning. In the previous experiments in Chapter 4, it has been demonstrated that the skull can be partitioned into 3 different orientations following 3 main axes (X,Y, and Z) and the results attested to the suitability of

such method. This chapter advances such approach by proposing a partitioning method which covers the geometrical aspects of the skull ensuring that the anatomical structure of the skull is well expressed. To achieve this, an improved skull registration method with faster computational speed is introduced and automatic partitioning is performed using fuzzy cmeans and label propagation. This produced regions or patches of the skull that are reminiscent of the conventional anatomical skull sites used in forensic morphological assessment. In order to assess the effectiveness of the proposed method, different local descriptors such as MeshLBP, Spin Image, and LD-SIFT are extracted on the local regions of the skull and dimensionality reduction techniques such as KPCA and KSVD are utilized to make the feature compact and distinctive. Eventually, predictive models such as KELM, SRC, and SVM are used to classify and determine the sex of each skull. The experimental results indicate that similar prediction results trend in Chapter 4 can be attained, where the best sex determination is 86% using Spin Image, KPCA, and SVM classifier. Besides, the effect of number of partitions has no statistical significance on the prediction results. Nevertheless, the results in this experiments are competitive with the results presented in forensic science, where a range of prediction between 70% - 90% has been reported in the literature (Garvin et al., 2014; Walker, 2008).

It can be concluded that computer vision approaches are suitable for sex determination with impressive generalization despite examining the possibility of using diverse approaches of skull partition, feature extraction, and classification techniques. In addition, the number of regions the skull can be partitioned does not have to be restricted, as it can be observed from results in Chapter 4 and 5 that number of regions has no significant impact sex determination results. This rather concludes the investigation on methods of skull partition, the effect of orientation of partition and number of partition on sex identification results, as well as the choice of feature extraction and classification techniques.

Future work can focus on the possibility of improving the registration method. Interest can be placed on techniques that operate very fast on large 3D point set, which also achieve good convergence without unnecessarily affecting the accuracy of registration. In another aspect, the technique for fusing two or more features can be improve further to accommodate other types of feature representation. Also, it can be noticed that the performance in Figure 5.7 seems to be oscillating. This can be attributed to some feature combination appealing better different number of skull partitions. Further investigation can be conducted on to gain better intuition by expanding the number of features considered for multiple feature combination.

Chapter 6

Novel Methods for Selecting Discriminant Anatomical Regions

> Discovery is seeing what everybody else has seen, and thinking what nobody else has thought
>
> Albert Szent-Gyorgi

This chapter presents the experiments conducted on sex estimation from human skull focusing on the selection of important skull regions. The proposed method in this chapter involves 3D skull preprocessing, skull partitioning into pseudo-anatomical regions, important skull region selection, 3D feature representation and classification. Similarly, the 3D preprocessing and feature extraction which have been discussed in Chapter 3 are used in this chapter.

6.1 Overview

In the process of determining the sex of skull, forensic experts also endeavor to find out which regions of the skull possess higher sexually dimorphic characteristics. In essence, this chapter is concerned with examining the human skull to discover the important local skull regions (sites) that contribute positively to the prediction results. This is crucial in gaining better intuition about factors influencing sex prediction. A particular common way to achieve this in the forensic domain is to attempt every possible combination of different dimorphic traits (Garvin, 2017). However, this sort of approach could quickly become exhaustive as the number of traits increases, because it involves a stepwise selection procedure where the feature selection algorithm either starts with no variables (features) in the model and proceed forward by adding one variable at a time or begins with all the potential variables in the model and proceed backward by removing one variable (feature) at a time. As such, this thesis proposes a novel method for learning important skull

regions directly from data using the concept of graph based group variable selection (Huang et al., 2012). The proposed method focuses on using the shape property of each local region to construct graphs which embed discriminating information about each local sub-region and use the graph as a form of regularization to displace unimportant skull regions. In summary, this chapter presents:

- The theoretical background of the proposed group variable selection.
- The techniques devise for learning selection of important skull regions automatically

The structure of the proposed method is illustrated in Figure 6.1. The process starts with 3D preprocessing to filter the input volumetric PMCT samples to remove noise, local irregularities, and isosurface (mesh) reconstruction to derive the hard tissue (skull layer). Afterward, the skull is partitioned into pseudo-anatomical regions by first registering the skull samples to a model skull using coherent point drift (CPD) to establish vertex point correspondence and then applying fuzzy c-means to cluster the skull into different regions. 3D local shape features are extracted on the partitioned regions, which are then placed in a skull region selection model based on group variable selection algorithm to discover the important skull regions. The resulting sparsely learned skull regions are used for training and testing different predictive models for sex determination as shown in Figure 6.1.

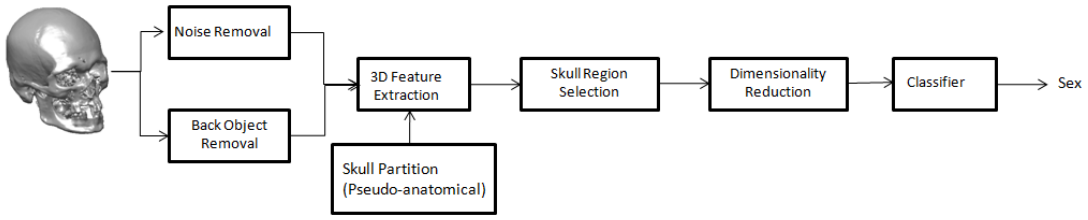


Figure 6.1: Structure of skull region selection and sex determination

6.2 Related Works on Skull Region Selection on Sex Identification

The interest of forensic experts is not only limited to achieving the predicted sex class of the skull, but to discover which skull regions contribute toward sex prediction. This section mainly discusses the previous works carried out on analysing the effect of combination of different parts of the skull on sex identification results and points out some literature on group variable selection.

6.2.1 Approaches to Skull Region Selection in Forensic Anthropology

Looking at the previous studies, some of the prominently assessed anatomical regions of the skull include the nuchal crest, mastoid process, supraorbital margin, glabella, mental eminence, and etc (Krüger et al., 2015; Walker, 2008). Walker (2008) examined 5 anatomical sites underunivariate and multivariate statistical analysis with linear and quadratic discrimanant analysis. It was discovered that under independent test, the glabella and mastoid process are more sexually dimorphic than other traits with prediction of 70%, with the nuchal crest showing the least dimorphic characteristics (57%).

Though the best result was achieved using the 5 traits with a prediction of 90%, a combination of only 3 traits produced a prediction of 88%. Graw et al. (1999) investigated how the shape of supraorbital margin can be assessed for sex determination and their result gave indication that this anatomical region can be independently used for sex prediction with an accuracy of 70%. Likewise, Pinto et al. (2016) studied this anatomical region based on computer-aided method by using 2D wavelet transform on 3D reconstruction of the scanned supra-orbital impression to study its shape variation. Williams and Rogers (2006) studied 21 sexually dimorphic traits for sex determination. Each trait was assessed and ranked according to its inter-observer error and accuracy when predicting male and female samples. The authors discovered that the size of the mastoid and supraorbital ridge, rugosity of the zygomatic extension, size and shape of the nasal aperture, and gonial anglewere ranked higher than the remaining characteristics, with 10% inter-observer error and accuracy 80%.

Langley et al. (2017) recently reported the analysis of the 5 traits of Walker and one additional trait (zygomatic extension) using decision tree model, with the aim eliminating the traits with high possibility of low inter-observer consistency. Conforming with previous standards, the authors adopted the standardized 5 ordinal-scale built by Buikstra and Ubelaker, which has been commonly reported in the literature. In order to make predictive, the ratings from these traits were used to construct classification and regression trees (CART). Their study showed that the decision tree which derived the combination of 3 traits (glabella, zygomatic extension, and mastoid process) provided the most accurate result with a prediction rate of 93.5 %, whereas the nuchal crest, supraorbital margin, and mental eminence were not chosen by the tree model.

With regard to morphometric analysis, Franklin et al. (2012) proposed a method of morphometric analysis, computing eight measurements of the 3D landmarks and

discriminant function analysis for prediction. Their study suggests high sexual dimorphism in the facial width (bizygomatic breadth) and the length and height of the cranial vault. The authors further used stepwise variable selection to discover the anatomical measures from 31 pre-selected landmark points that contribute significantly to sex determination on Western Australian samples (Franklin et al., 2013). Similarly, AbdelFatah et al. (2013) conducted 19 measurements on the cranial via annotation of the anatomical points and use stepwise variable selection to find the important cranial traits.

Bigoni et al. (2010) examined 82 ecto-cranial lamdmarks from 139 cranial samples of the Central European population. The landmarks originate from 7 subregions of the skull such as the configuration of the neuro- cranium, cranial base, midsagittal curve of vault, upper face, orbital region, nasal region, and palatal region. Using Generalized procrustes analysis (GPA) for analyzing the shape configuration, the authors noticed no sex difference in the sample set when landmarks from the whole skull were used. However, through partial shape examinations of each of the seven regions, there were indications of strong sexual dimorphism in the mid-sagittal curve, the upper face, the orbital region, the nasal region, and the palatal region, but no sex variation in the cranium base and the neurocranium configuration.

Another study has assessed two discriminant function analysis methods on 17 craniometric variables from 90 Iberian skulls (Jiménez-Arenas and Esquivel, 2013). The authors observed high sexual dimorphism on the Bizygomatic breadth with 88.5% prediction rate.

Though the previous works have shown that some skull regions possess better sexually dimorphic characteristics than others, the approach of deriving the importance of traits is significantly based on the prior knowledge of the forensic expert in localizing those skull regions and the discovery of highly dimorphic trait is achieved by simply conducting step-wise combination of different traits in a DFA model. This traditional approach of discovering the importance of skull regions can become less efficient in a situation where several skull regions are considered, since the computational cost and time will increase dramatically. Thus, improved methods which will provide robust results with less computational complexity are necessary.

6.2.2 Group Variable Selection

The aim of this section is to discuss some of the advancement in the domain of group variable selection and their applications.

In a standard regression problem, one would be interested in finding the coefficients or variables of the model that fit the data to a linear function and in most cases, it is expressed as:

$$y = Xw + \varepsilon \tag{6.1}$$

where y is a vector, w is a vector of coefficients, ε is the error term, and X is the data matrix. Imposing constraint such as l_1 norm on the function leads to penalization of some of the coefficients, which makes them sparse and example of such algorithm is least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996). However, in some cases the data at hand has a structural property which must be preserved. As such, a direct application of LASSO on the data will lead to sub-optimal solution and also ignore the spatial configuration of the data. This has resulted in the introduction of the concept of group variable selection, which does not only maintain the structure of the data, but lends the data well to interpretation (Huang et al., 2012).

Bakin. (1999) proposed group LASSO with its computational properties. Yuan and Lin (2006) developed group LASSO by splitting the original data into disjointed groups and imposing Euclidean norm on each group which leads to selection of groups rather than individual variables. Meier et al. (2008) extended the concept of group LASSO for logistic regression and demonstrated its application to the problem of learning from high dimensional data. Zhao et al (2008) proposed a group variable selection method that uses composite absolute penalty, which incorporates side information to the fitting of the regression model. There are also studies focusing on constructing structured sparsity through overlapping groups (Jacob et al., 2009). Rao et al. (2016) proposed a much less restrictive method of structured feature selection, by considering the noting overlap between groups has a form of loose similarity, which primarily captures the prior knowledge about the data. Thus, in a situation where some features from a particular have weak representation power, this can be accounted for by other groups, since they share some form of similarity and overlap. In order to preserve the structural locality of the groups, some authors have proposed graph-based regularization for group LASSO (Chen et al., 2011; Jacob et al., 2009).

Zhu et al (2013) introduced a group variable selection method for learning representations for video-tagging. The approach uses the temporal-spatial prior knowledge gather from constructing laplacian graphs from the video data, then imposes inter-group and the intra-group sparsity. Yang et al. (2017) proposed two-level structured sparse model referred to as Sparse Group Lasso with Group-level

Graph structure (SGLGG) for Genome-wide association studies (GWAS), which is an integration of sparse group LASSO with group level group LASSO. Most of the previous works on group variable selection are concentrated on addressing overlapping groups and forming of graphs which maintain the structure of the data. Chapter 6.3.2 introduces techniques for constructing discriminant graphs which maximize the variation between the local structure of the data, whilst preserving the within group structure.

6.3 Region Selection

This section describes the group variable selection algorithms used for determining the important skull regions that contribute positively to sex determination results.

6.3.1 Group Lasso

Lasso is a regression analysis method that performs two tasks at the time: (1) variable selection (2) regularization in order to improve generalization result of a model and also provides good interpretation of the model. Lasso is mainly based on solving a least square problem of:

$$\min_{w \in \mathbb{R}^p} \left(\frac{1}{N} ||y - Xw||_2^2 \right) \text{ subject to } \; ||w||_1 \leq \lambda \tag{6.2}$$

where X is a $M \times p$ matrix of input data, y is a vector of M observations, w are coefficients of the model, and λ is a regularization value. Due to the standard constraints imposed on Lasso: $||w||_1 \leq \lambda$, the l_1 norm can induce sparsity on w to obtain some coefficients that are exactly equal to zero. This makes lasso useful for variable selection, as those coefficients can be displaced from the model.

In some regression problems, the interest may be on selecting a group of variables or covariates that possess common data or feature properties. Such problem can not be directly tackled by the standard Lasso. Thus, an improved formulation, Group Lasso, has been proposed to select group variables or covariates (Friedman et al., 2010). Basically, there are two perspectives to this problem: (1) is to induce sparsity on a group by setting the entire group to zero if the sum of euclidean norm of a set of coefficients in the group is ≤ 0 (Yuan and Lin, 2006). (2) is to induce a within and between group sparsity, so that when a set of coefficients in a group are zero, only those coefficients will be displaced. This is different to the standard Lasso because the within group sparsity depends on the structural property of the group (Friedman et al., 2010).

In this work, due to the manner in which the data is structured, the first approach initially introduced by Yuan and Lin (2006) is adopted. More clarification on the structure of training data will be discussed in the next section.

Now assuming that the p predictors can be divided into K groups, and p_k denoting the predictors from group k. Hence, the corresponding data matrix and coefficients for that group will be X_k and w_k respectively. The problem can be formulated as convex optimization problem in the form (Friedman et al., 2010; Yuan and Lin, 2006):

$$\min_{w \in \mathbb{R}^p} \left(||y - \sum_{k=1}^{K} X_k w_k||_2^2 + \lambda \sum_{k=1}^{K} \sqrt{p_k} ||w_k||_2 \right) \tag{6.3}$$

where $\sqrt{p_k}$ is the term that accounts for the size of the group and $|| \cdot ||$ is the euclidean norm (non-squared).

To compute the group lasso, each sub-problem will take the sub-gradient form of (Friedman et al., 2010; Yuan and Lin, 2006):

$$-X_k^T \left(y - \sum_k X_k w_k \right) + \lambda \cdot s_k = 0; \quad k = 0, 1, 2, \ldots K, \tag{6.4}$$

where $s_k = \frac{w_k}{||w_k||}$ if $w_k \neq 0$ and it is a vector with $||s_k||_2 < 1$ otherwise (Friedman et al., 2010; Yuan and Lin, 2006). Assuming the solutions of the groups are $\hat{w}_1, \hat{w}_2, \ldots \hat{w}_k$. If

$$||X_k^T \left(y - \sum_k X_k \hat{w}_k \right) || < \lambda \tag{6.5}$$

then $\hat{w}_k$ is equal to zero; otherwise it satisfies (Friedman et al., 2010; Yuan and Lin, 2006):

$$\hat{w}_k = \left(X_k^T X_k + \frac{\lambda}{||\hat{w}_k||} \right)^{-1} X_k^T r_k \tag{6.6}$$

where

$$r_k = y - \sum_{k \neq l} X_k \hat{w}_k \tag{6.7}$$

The solution provided in (6.6) can be simplified as $\hat{w}_k = (1 - \frac{\lambda}{||v_k||}) v_k$ assuming $X_k^T X_k = \mathbf{I}$ and $v_k = X^T r_k$.

Though this form of group variable selection is able to discover the useful groups and eliminate the unimportant ones, its main limitation is the inability to effectively handle data requiring preservation of structure of its group properties, for instance the locality of the skull regions. Moreover, the standard group LASSO algorithm is unable to handle data which share some element of overlap between

groups (Huang et al., 2012). As discussed in Bach. (2008); Wang and Shizhong (2009), group LASSO only satisfies selection consistency under low correlation between data. Meaning that if there is low correlation between the groups of the data, it is guaranteed that group LASSO will exhibit consistency in selection of group variables. Otherwise there can be no consistency. This is the main reason why a new group regression model for manifold data, which is able to preserve such property is proposed in this Chapter.

6.3.2 Novel Group Regression Model for Manifold Data

The aim of the proposed algorithm is to mitigate the limitation of standard group LASSO algorithm (such as inability to preserve the structure of group data). This is achieved by using the concept of manifold data representation to obtain the secondary shape property of each local skull region as graphs and formulating a method of information induction through division of the graph computed for a skull region by the graph of the entire skull regions. This approach makes the graph of that particular skull region discriminant and distinctive when solving for the group coefficients, which eventually helps in preserving the structure of the partitions of the skull. In order to achieve the proposed algorithm, graph is constructed for each skull region, then the eigendecomposition of the graphs is learned, and finally information induction is computed for each graph to make the graphs discriminant. The derived discriminant graphs are then placed in group regression model objective to determine the group coefficients.

Graph based method is a form of regularization for learning the relationship between data, by constraining the function learned on the data with the shape of the data itself. The main idea is such that only a few subspace of the data contains the meaningful information that need to be learned. This information is said to exist in a manifold. Given a sample set X, the standard manifold learning constructs an undirected graph $G = (V, E)$, with the vertex set $V = X$ and W_{ij} is the edge weights in the adjacency graph, which is defined by $W(x, x') = \exp\{-||x - x'||^2/2\sigma^2\}$. The unnormalized graph Laplacian of G is defined as $L = D - W$, where D is a diagonal matrix with diagonal entries $D_{ii} = \sum_{j=1}^{l+u} = W_{ij}$. Considering that the eigenmaps can be obtained as:

$$\varphi(X_k) = XA \tag{6.8}$$

where A consists of the eigenvectors $\mathbf{a}$ of a generalized eigenvalue decomposition problem $XLX^T\mathbf{a} = \lambda X^T DX\mathbf{a}$. This research proposes to embed the manifold in group regression model and re-formulate objective as:

$$\min_{w \in \mathbb{R}^p} \left(||y - \sum_{k=1}^{K} \varphi(X_k) w_k||_2^2 + \lambda \sum_{k=1}^{K} \sqrt{p_k} ||w_k||_2 \right) \tag{6.9}$$

where $\varphi(X_k)$ is now the manifold defined on each group data. Note that to obtain $\varphi(X_k)$ the corresponding L and D for deriving A in Equation 6.8 are derived based on X_k only. For the sake of clarity, the essence of constructing graph in this thesis is to exploit the secondary information about the cranial data, and computing such graphs for each independent region provides the possibility of having varying graph shape information of the regions. This is because the values of eigendecomposition solution A of each region's graph is different, so this information is informative for the proposed group variable selection algorithm to learn the important regions while keeping the locations of the regions. This is contrary to simply constructing a single graph for the entire skull, as it will offer no variation among the regions since the eigenvalues are the same, and its application to the group regression simply offers no difference to using the original data X, which is similar to applying a constant to the original data.

Each independent group has a graph describing the relationship among data in that group. By learning such formulation, a different solution is gained where group variables are selected instead of individual variables by penalizing the coefficients of the manifold by an Euclidean norm (squared) $\sum_{k=1}^{K} \sqrt{p_k} ||w_k||_2$. However, it is worth noting that such notion does not necessarily have a holistic influence on the group data, since each group will only be influenced by the graph from its group. Therefore, this thesis proposes to construct the graph from a particular group as an induction of every graphs of the remaining groups. To achieve this, the graph of a particular group is divided by the summation of the graphs from other groups as:

$$\mathbf{D}_k = \left\{ \sum_{k \neq l}^{K} \varphi(X_k) \right\}^{-1} * \ \varphi(X_k) \tag{6.10}$$

This serves as an induction of the main group and the rest of the groups, which is representative of the relation of within and between group variables. This is a way of capturing and contrasting the difference between the manifold of a particular sub region and the overall manifold information of the rest of the sub regions. Such difference for each sub-region is then preserved through the optimization with respect to the coeffecients w_k in Eq. 6.11. This eventually makes the resulting Laplacian graph discriminant.

Thus the new formulation can be expressed as:

$$\min_{w \in \mathbb{R}^p} \left(||y - \sum_{k=1}^{K} \mathbf{D}_k w_k||_2^2 + \lambda \sum_{k=1}^{K} \sqrt{p_k} ||w_k||_2 \right) \tag{6.11}$$

Different to the standard group LASSO in Eqs 6.4 - 6.7 where the residual is used in the analytical solution to the objective function, the solution for each group matrix in the proposed algorithm is obtained in form of least square regression as:

$$\left(y - \sum_{k} \mathbf{D}_k w_k \right) + \lambda \cdot s_k = 0; \quad k = 0, 1, 2, \ldots K, \tag{6.12}$$

where the solution of the group coefficient w_k is obtained analytically as:

$$\hat{w}_k = \left(\mathbf{D}_k^\top \mathbf{D}_k \right)^{-1} \mathbf{D}_k^\top y \tag{6.13}$$

and $s_k = \frac{w_k}{||w_k||}$ if $w_k \neq 0$ and it is a vector with $||s_k||_2 < 1$ otherwise similar to the criteria used in (Friedman et al., 2010; Yuan and Lin, 2006). Since, there is already an induction graph obtained in the form of $\mathbf{D}_k$ which induces information between the graph of a region and the remaining graphs (meaning that a graph partially carries information about other graphs), thus each group matrix is treated as a regression problem where the true label y will depend linearly on the k graph variables. Hence, the need for residual is not necessary because residual on the other hand will offer the linear relationship between graph and the corresponding predicted values.

Implementation

To model the proposed group variable selection algorithm, training data is mapped into the basis space to obtain their new representations $\varphi(X_k)$ based on the concept of manifold learning. The **kNN** is used to generate the graph and in particular the heat kernel is used to generate the k nearest neighbor. By generating $\varphi(X_k)$, the discriminant induction graph $\mathbf{D}_k$ is obtained using Equation 6.10. The algorithm of the proposed is illustrated as follows:

Algorithm 3 Novel Group Regression Model for Manifold Data

Input: $X = [X_1, X_2, \ldots, X_k]$
\\ *Make a graph for each group offline*
Construct graph L on each X_k
Compute $\varphi(X_k)$ following Eq. 6.8
\\ *obtain the discriminant graph*
Compute discriminant graph $\mathbf{D}_k$ using Eq. 6.10
\\ *run Group Regression on constructed graphs*
while W not converged **do**
  for $k = 1, \ldots K$ **do**
    Obtain w_k using Eq 6.11
  end for
  $W = [w_1, w_2, \ldots, w_k]$
end while
Output: W

The formulation is convex optimization problem, which can be solved following (Yuan and Lin, 2006). This sort of approach has been described with proof by Zhu et al (2013) when implementing group lasso for video tagging application. However, the proposed approach is different since it considers constructing an induction discrminant graph as described in Equation 6.10.

6.3.3 Novel Group Regression Model for Manifold Data with Split Augmented Lagrangian Shrinkage Algorithm (SALSA)

To improve the performance of the algorithm introduced in Section 6.3.2, this thesis proposes to use an optimization solver: split augmented Lagrangian shrinkage algorithm (SALSA) (Afonso et al., 2011) based on the concept of augmented langragian to solve Equation 6.11. The optimizer combines both augmented lagragian and variable splitting in an iterative algorithm which provides a more flexible way of handling the problem.

SALSA is an optimizer which has been used for image inverse problem such as restoration and reconstruction. Of recent, its application to general least square minimization was demonstrated in (Selesnick, 2011). In this research, SALSA is initially used to solve the inner least square problem of each group (i.e the least square which target the coefficients (column space) instead of the groups) before applying the euclidean norm to determine the penalization of the groups. Assuming the minimization problem of the inner least square of each group k can be expressed as (Selesnick, 2011):

$$\arg\min_{w_k} \frac{1}{2}||y - \mathbf{D}_k w_k||_2^2 + \ ||\lambda \odot w_k||_1 \tag{6.14}$$

where the $\lambda \odot w_k$ is the element-wise multiplication of the equal size vectors, λ and w_k. When all elements are equal then it simply becomes w_k. However, in the idea of variable splitting, it is commonplace to create a new term which allows for non-uniform regularization. Thus the variable splitting problem can be expressed as:

$$\arg\min_{w_k,u_k} \frac{1}{2}||y - \mathbf{D}_k w_k||_2^2 + \ ||\lambda \odot u_k||_1 \tag{6.15}$$

$$\text{such that} \ \ u - w = 0 \tag{6.16}$$

The problem in Eq 6.16 can be expressed in the augmented lagrangian form. Briefly recall that for a constrained optimization problem:

$$\arg\min_{z_k} E(z) \qquad \text{subject to} \ \ \mathbf{D}_k \text{z - y} = 0 \tag{6.17}$$

where b $\in \mathrm{R}^P$ and $\mathbf{D} \in \mathrm{R}^{\mathrm{pxn}}$ meaning that there are p linear equality constraints. The augmented Lagrangian of the problem can be expressed as:

$$L(z, \lambda, \mu) = E(z) + \lambda^T(\text{y} - \mathbf{D}_k z) + \frac{\mu}{2}||\mathbf{D}_k z - \text{y}||_2^2 \tag{6.18}$$

where $\lambda \in R^P$ is a vector of Lagrange multipliers and $\mu \geq 0$ is penalty parameter. The Augmented Langragian Method simply iterates between minimizing $L(z, \lambda, \mu)$ with respect to z, keeping λ fixed, and updating λ until some convergence criterion is satisfied.

The problem in Eq 6.16 can be expressed in the form of Eq. 6.17 by setting:

$$z_1 = w_k, \ z_2 = u_k, \ z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{6.19}$$

$$E(z) = \frac{1}{2}||y - \mathbf{D}_k z_1||_2^2 + ||\lambda \odot z_2||_1 \tag{6.20}$$

Then, the iterative augmented lagragian method (ALM), also known as method of multipliers (MM) (Hestenes, 1969; Powell, 1967) can be applied as (Afonso et al., 2011)

Algorithm 4 ALM/MM

Initialize $\mu > 0, d_k$
repeat

$$w_k, u_k = \arg\min \frac{1}{2}||y - \mathbf{D}_k z_1||_2^2 + ||\lambda \odot u_k||_1 + \frac{\mu}{2}||u_k - w_k - d_k||_2^2 \tag{6.21}$$

$$d_k \leftarrow d_k - (u_k - w_k) \tag{6.22}$$

end when stopping criteria is reached

where d_k is the estimate coefficient which should be initialized prior to iterative process. In this work, d_k is initialized to zero. In order to attain the global minimum, the alternating direction method of multipliers (ADMM) (Boyd et al., 2011) is used to alternate between w_k and u_k, which can be expressed in **Algorithm 5**.

Algorithm 5 ADMM

Initialize $\mu > 0, d_k$
repeat

$$u_k \leftarrow \arg\min_{u_k} ||\lambda \odot u_k||_1 + \frac{\mu}{2}||u_k - w_k - d_k||_2^2 \tag{6.23}$$

$$w_k \leftarrow \arg\min_{w_k} \frac{1}{2}||y - \mathbf{D}_k z_1||_2^2 + \frac{\mu}{2}||u_k - w_k - d_k||_2^2 \tag{6.24}$$

$$d_k \leftarrow d_k - (u_k - w_k) \tag{6.25}$$

end when stopping criteria is reached

Once the coefficient estimate d_k has been obtained, they are used as the vector for each group and the group sparsity is determined if $\frac{w_k}{||w_k||} = 0$ otherwise the group is not sparse because $\frac{w_k}{||w_k||} < 1$ but not 0. The overall algorithm is illustrated in **Algorithm 6**

Algorithm 6 Novel Group Regression Model on Manifold Data with SALSA

Input: $X = [X_1, X_2, \ldots, X_k]$
\\ *Make a graph for each group offline*
Construct graph L on each X_k
Compute $\varphi(X_k)$ following Eq. 6.8
\\ *obtain the discriminant graph*
Compute discriminant graph $\mathbf{D}_k$ using Eq. 6.10
\\ *run SALSA to solve Group regression on constructed graphs*
Initialize $w_k = w_0$
for each k **do**
    Apply Eq. 6.16
    Apply **Algorithm 4** and **Algorithm 5**
    Set $w_k = d_k$
    Compute $\frac{w_k}{||w_k||}$
    $W = [w_1, w_2, \ldots, w_k]$
end for
Output: W

6.3.4 Multi-scale Region Learning/Selection

In Chapter 5, each number of partitions (3 to 99) from the pseudo-anatomical partitions are evaluated independently to observe the influence of those partitions on sex prediction results. To achieve automatic selection of regions that contribute to prediction, over-representation is performed by repeatedly partitioning the skull from coarse to fine scale. The objective is to select a set of partitions (sub-regions) from a skull data, with indication that those selected partitions have particular influence on the prediction results.

Assuming there are indexes of mesh points, a design (data) matrix $X \in R^m$ of disjointed groups $\{L_1, L_2, \ldots L_K\}$ can formed. Thus, the design matrix is represented as $X = (X_{L_1}, X_{L_2}, \ldots X_{L_K})$, where the features from each group are: $X_{L_K} = X_1, X_2, \ldots X_{nk}$. The groups are obtained by using the pseudo-anatomical partitioning in Section 5.3.2. Therefore the interest in this section is to derived a set of group co-efficients where the sparse values is equivalent to the regions that should be displaced from the skulls.

This is performed in 5 different steps:

1. Design data: a total of 50 random sample skulls are chosen to model the group regression algorithm.

2. Deriving sub-regions: in this case, each skull in the design data is partitioned into multi-scale of $[5, 11, 15, 17, 19, 21]$ regions, resulting in a total of 88 sub-regions (groups) and 50×88 matrix. As already described in Section 5.3,

again the correspondence already established has been exploited to ensure that the sub-regions are properly propagated across the training data. For each group, the local descriptor is computed and aggregated in similar fashion to the one in Section 5.3. Therefore, the total dimension of features from each skull is (320×88 sub-regions $= 28160$ features). These features from different skulls form the design matrix of 50×28160, which group regression is used to induced group sparsity.

3. Defining the group structure Z: the structure of the groups are calculated as: $Z = \frac{\text{number of sub-regions}}{S}$ where S is the number of structures to divide the training matrix. In this experiment 88 sub-regions are used. Meaning that the design matrix X is divided into 88 different structures each containing a partition of the skull.

4. Run the group regression model: for a particular value of regularization value (from $2^{-15} : 2^{10}$) the sparse groups can be computed by using Eq. 6.3, following the mathematical derivation in section 6.3.1 for group LASSO, and solutions in Sections 6.3.2 and 6.3.3 for the proposed group variable selection method and its SALSA based optimization.

5. Selecting sparse regions: assuming the total coefficients from group regression is $\hat{W} = \hat{w}_1, \hat{w}_2, \ldots \hat{w}_k$. The sparse group of a new sample x_i can be select by first following steps 2 to 3, before computing the hadamard (element-wise) product as:

$$\hat{w}^* = x_i \odot \hat{W} \tag{6.26}$$

where $\hat{w}^*$ denotes the sparse index of the new sample.

6.4 Sex Determination

In order to predict the sex of the skulls, the same 100 skull samples (with 60% training and 40% testing) are used in this experiment. Each sample skull is run through the sparse group coefficients as explained in step 5 above, results in skulls with sparse patches (regions). The non-sparse patches (regions) are collected for feature representation using 3D local shape features such as MeshLBP, spin image, and LD-SIFT. The resulting features are aggregated and concatenated as described in Section 5.3. Afterward, the features are processed with dimensionality reduction algorithm such as KPCA and KSVD, and finally classified using SVM, KELM, and SRC for sex determination. The results obtained from the experiments are illustrated in Figures 6.3, 6.4 and Table 6.1.

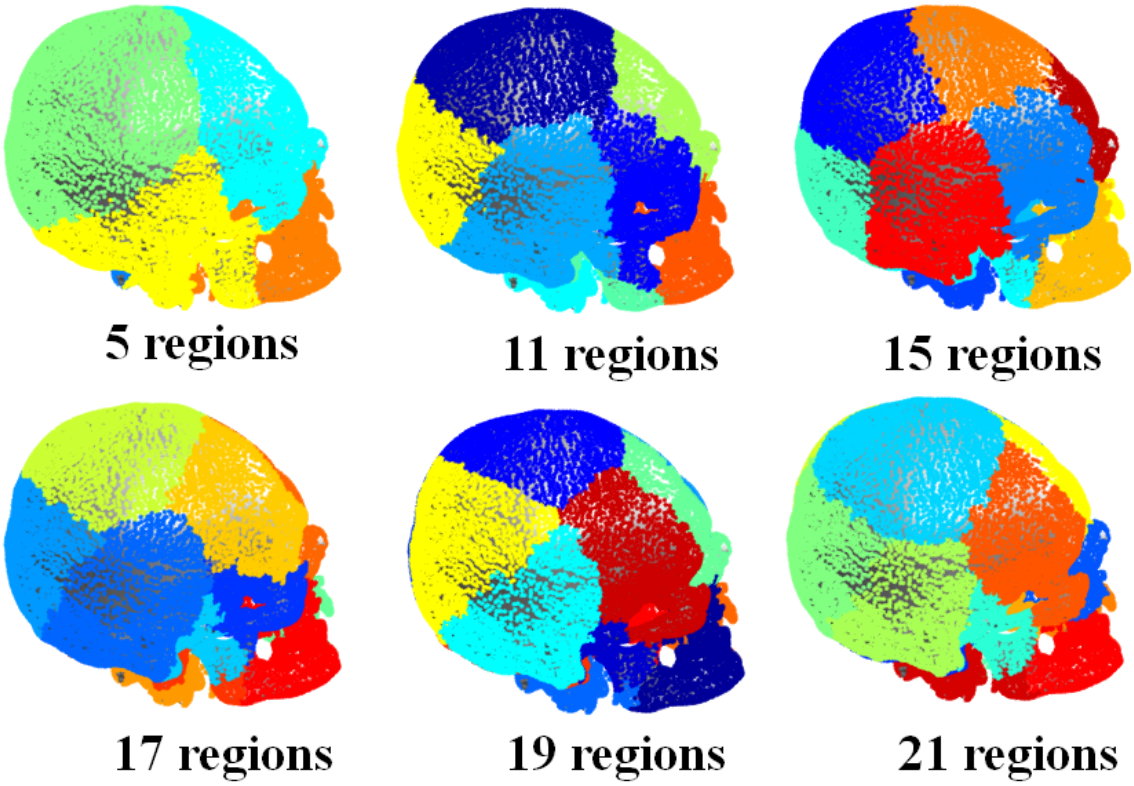


Figure 6.2: Illustration of pseudo-anatomical partitioning with FCM

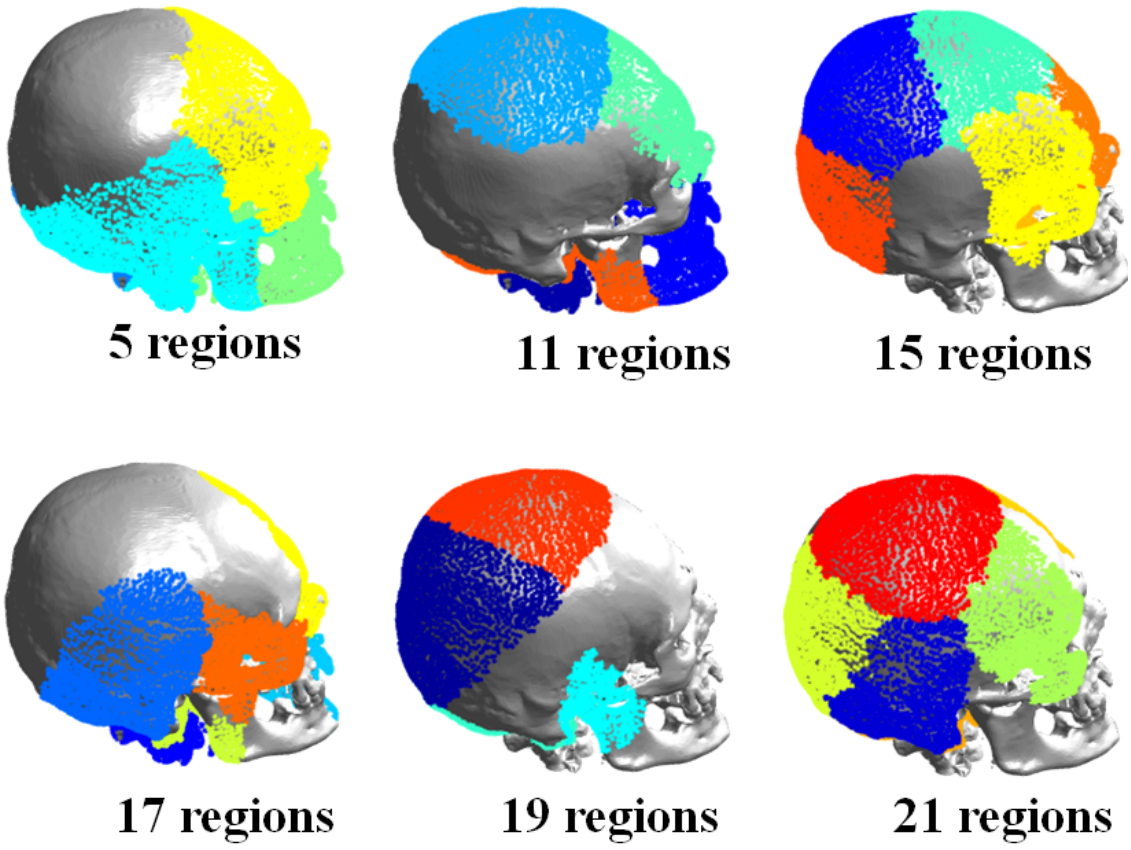


Figure 6.3: Illustration of group patch selection from the clustered regions by group LASSO

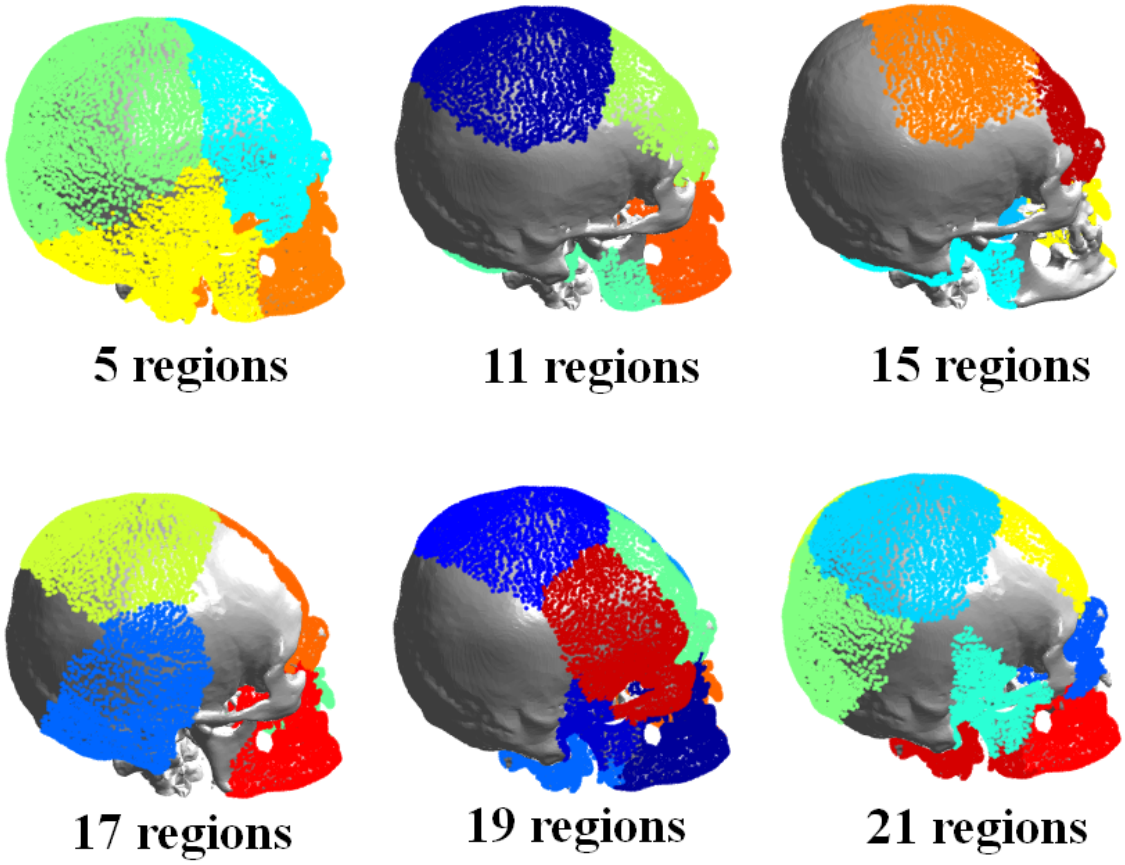


Figure 6.4: Illustration of group patch selection from the clustered regions using the proposed methods

This procedure for discovering the important regions based on group variable selection has been repeated for different values of regularization, C but the best value in this experiment is 2^{-6}. The results in this experiment are promising both from the perspective of discovering important skull regions and considering them for sex determination of 3D skulls.

Table 6.1: Result of skull region selection based on group variable selection

Feature Extraction	Classifier	Group LASSO (%)	Graph Group based Regression (%)	Graph Group based Regression with SALSA (%)
LD-SIFT	KELM	78.6	84	84
	SVM	79.7	**84.5**	83
	SRC	77.6	73	78.9
Spin Image	KELM	82.4	81	82.2
	SVM	83.3	81	83.5
	SRC	**83.7**	71.5	80.8
MeshLBP	KELM	81	78.8	77.7
	SVM	**82**	78.4	80
	SRC	81.7	81	81.8
LD-SIFT + KSVD	KELM	78.3	79	78.3
	SVM	80.4	78	72
	SRC	77.4	**81**	75.6
Spin Image + KSVD	KELM	78	**81.2**	80.6
	SVM	79	80	80.6
	SRC	79.8	81	81.2
MeshLBP + KSVD	KELM	75.3	75.7	75
	SVM	76	75.6	70
	SRC	74	71.2	**79.6**
LD-SIFT + KPCA	KELM	80.4	79	80
	SVM	79.3	82	**83.8**
	SRC	81.9	81	79.2
Spin Image + KPCA	KELM	82.9	81.6	80.5
	SVM	83.75	**84.8**	84.5
	SRC	82.8	82	82
MeshLBP + KPCA	KELM	76.5	80.5	81
	SVM	81	**84.4**	83.3
	SRC	73.5	75	77

The results depicted in Figures 6.3, 6.4 represent the output of the two group variable selection algorithms. Each skull in each figure depicts the important regions selected at a particular level (for instance 11 regions) after modelling the group variable selection algorithms. Intuitively, it can be observed that the manifold (graph) construction is influential in preserving the locality of the partitioned regions and also in guiding rotational invariant descriptors such as MeshLBP. To decipher the importance of graphs on learning group variable selection, the regions automatically chosen from each scale are examined as shown in Figure 6.4. It can be noticed that the selected regions are aligning along similar geometrical locality of skull, while in Figure 6.3, it can be noticed that the regions are selected inconsistently. This is quite interesting because each scale with its regions has been learned independent of regions from other scales.

It can also be noticed that some regions are selected (or not selected) at coarser scale and vice versa at other scale, especially in the result of standard group LASSO. This is attributed to the fact that as the number of partitions increase, the regions become more disjoint and independent making the algorithm to learn such smaller regions as individual selection problem. Moreover, due to the share overlap between some of the features in the two scales, the group LASSO algorithm basi-

cally displaced the entire region without differentiating information of the smaller regions to the large one. Nonetheless, it can be observed that the graph based group regression model is still able to preserve the locations the smaller regions.

In addition to the visual assessment, a simple analysis is conducted to measure the correlation coefficients between 2 different skull regions. Meaning that from two different number of regions, 2 random sub-regions are selected and their correlation coefficient is measured. For instance, if 5 regions and 11 regions are considered as the two different number of skull regions, a sub-region X_2 is selected from 5 regions, while another sub-region X_{12} is selected from 11 regions, and their correlation coefficient is computed. The results show that the correlation among regions in the original data (used for group LASSO) is higher than when manifold (for graph group regression) is used to represent the regions. Table 6.2 illustrates the results of the correlation check.

Table 6.2: Results of measuring the correlation coefficient of two different skull regions

Number of Regions	Regions (patches)	Group LASSO	Graph Group based Regression
5 regions = $X_1, \ldots, X_5$	X_1 vs X_3	0.9831	0.1547
11 regions = $X_6, \ldots, X_{16}$	X_2 vs X_{12}	0.9483	0.055
15 regions = $X_{17}, \ldots, X_{31}$	X_8 vs X_{15}	0.9851	0.423
17 regions = $X_{32}, \ldots, X_{48}$	X_{26} vs X_{45}	0.9850	0.2591
	X_{14} vs X_{48}	0.9772	0.2258

With regard to sex determination from the sparse pseudo-anatomical regions, the best result 84.8% is within the commonly reported range in forensics. It is worth noting that sex determination is only performed after the important regions have been selected, in order to their influence on prediction performance. The best result is attained extracting spin image from the sparse regions discovered through the proposed group variable selection and applying KPCA and SVM on the features for classification.

6.5 Conclusion

This chapter investigated how adequate selection of important skull regions influences sex determination results. Two novel algorithms: (1) Novel Group Regression Model for Manifold Data and (2) Novel Group Regression Model for Manifold Data with Split Augmented Lagrangian Algorithm, which are based on the concept of group variable selection are proposed to automatically select the important skull regions by directly learning from the skull data. The two approaches make use of a

strategy of imposing group regularization on the group coefficients learned directly from skull regions which have been represented in form of manifold using the concept of graph based methods. The main contribution of the proposed methods is to mitigate the limitations of group LASSO such as inconsistency which results in inability to preserve the structure of data. Thus, this thesis proposed to derive another representation for the skull regions (groups) by computing their manifold representation, which is believed to offer a different kind of information than the original data. Then, in order to make the manifold of each group distinctive (discriminant), manifold induction is computed which mainly divides the manifold of a group by the manifold of remaining groups. The objective here is to form a within-group/between-group representation. It is believed that such representation will reduce correlation between the groups, and through solving for the group coefficient w_k, the algorithm can find patterns that are more consistent than when the correlation between group is very high which is experienced in standard group LASSO. Compared to the standard group variable selection (group LASSO) the experimental results indicate that the proposed region selection methods which incorporate graph construction provide better consistency in the discovered local regions than those without graph.

Chapter 7

Validation Study

> The aim here is not to separate fact from fantasy but to show how each embodies a distinct class of knowledge and how one is deeply implicated in the other.
>
> Constance Penley

The preceding chapters have demonstrated the suitability of using CV based methods for solving sex identification problem on 3D skull data. This chapter serves as a validation of the studies presented in the previous chapters by conducting 3 main analyses which are data split evaluation, comparison of CV based solutions with traditional forensic techniques, and assessment of the influence of population groups on sex identification performance.

7.1 Effect of Data Splitting on Sex Determination

The aim of this section is to determine whether the choice of training and testing set split affect prediction result. To achieve this, different methods of data splitting methods are considered:

- Leave one out: a particular sample is held out for testing while the remaining samples are used for training, which is then repeated across all samples and the average prediction result is calculated.
- k-fold cross validation ($k = 10$): involves dividing the entire dataset into k segments. A particular segment is held out for testing, while the remaining segments are used for training. This is repeated across all segments and the average prediction result is calculated. In this case, the dataset is partitioned into 10 segments, a segment is selected for testing while 9 segments are used for training and each segment taking a turn in the testing set.

- Also, additional random splits are evaluated in this thesis for sex identification, which are:
  - 50 training and 50 testing samples
  - 70 training and 30 testing samples

The results of these experiments are presented in Table 7.1

Table 7.1: Performance comparison based on data splitting method

Data split technique	Prediction rate (%)
Leave one out	88
k-fold cross validation	87.4
Random (50/50)	84
Random (70/30)	87

The reason for these experiments is to justify why random splitting method of 60% training and 40% testing has been considered in all experiments in this thesis. It can be seen from Table 7.1 that other data splits do not significantly influence the prediction results. Yet, it should be noted that the random split considered in this research is a fairer technique of data split which creates a balance between training and testing samples. First, it should be noted that in the case of 60/40 split, the samples used for training are completely different to the testing set. However, in leave one out and k-fold all data samples would have in one way or another been used in turn for training and testing, thus no independent test is performed with an unused samples. Second, even though a separate 5-fold cross validation is performed in 60/40 split during training to determine the best regularization parameter to fit the model, the amount of training samples is not critically depleted. This is unlike the 70/30 and 50/50 splits which have similarities to 60/40 since 5-fold cross validation is performed during training and separate test is conducted using the unused test samples. However, for 70/30 split the unused test sample is too small to gain true reflection of model performance, while 50/50 split will eventually result in lesser training samples.

7.2 Effect of Population Group on Sex Determination

The aim of this section is to evaluate whether pooling samples from a particular population group for training a classifier and testing with samples from entirely different group will have an effect on the prediction results. This is performed inline with the approach which have been discussed in the literature, where a number of related works have performed similar analysis (Garvin et al., 2014; Guyomarc'h

and Bruzek, 2011). In this research, the total number of Malaysian samples is 71, while Non-Malaysian samples is 29. Therefore, the feature representation and classification technique based on MeshLBP, KPCA, and SVM is evaluated since this combination has produced the best prediction results under the two partitioning methods. The results are presented in Table 7.2

Table 7.2: Performance comparison based on population group

Estimation Method	Training	Testing	Prediction rate (%)
MeshLBP & KPCA & SVM	Malaysian	Non-malaysian	80
MeshLBP & KPCA & SVM	Non-malaysian	Malaysian	76.1

From the results, it can be concluded that the proposed approach is not affected by difference in population groups since the result generalized to the two groups. Using Malaysian samples as training set and Non-Malaysian as testing the result is 80% which is close to the best result in this research (86%). In the case where Non-Malaysian samples are used for training and Malaysian samples for testing, it can be seen that the prediction rate is 76%, however it is worth noting that the training sample (29) is much less than the testing sample (71).

7.3 Comparison with Forensic Approach

This section presents the performance comparison of the proposed CV based techniques for sex determination with conventional forensic methods.

7.3.1 Morphometric Analysis

For the sake of completeness, the performance of the proposed CV methods is compared with conventional forensic estimation method (Ibrahim et al., 2017) as shown in Table 7.3. The forensic method (Ibrahim et al., 2017) is based on traditional morphometrics, where 22 estimation parameters are measured covering anatomical locations of 87 cranial samples (45 males and 42 females). These 87 samples are drawn the same dataset obtained from Hospital Kuala Lumpur (HKL). The measurements include the Maximum cranial length, Maximum cranial breadth, Cranial base length, Basion-bregma height, Bizygomatic breadth, Maximum frontal breadth, Minimum frontal breadth, Basion prosthion, Upper facial height, Nasal height, Nasal breadth, Orbital height of the right eye, Orbital height of the left eye, Orbital breadth of the right eye, Orbital breadth of the left eye, Biorbital breadth, Auriculo-bregmatic height of the right side, Auriculo-bregmatic height of the left side, Naso-occipital length, Mastoidal bregma height of the right side, Mastoidal bregma height of the left side, Nasion bregma height. Through independent t-test,

the authors were able to deduce the difference between male and female categories. Discriminant function analysis (DFA) is then used to generate equations for modeling the measurements and regression analysis to classify the samples into their respective classes. The prediction result from the study obtained through cross validation is between 78.2 - 86.2%. The proposed method in this thesis attained 86% on heuristically partitioned skull and 86.5% on pseudo-anatomically partitioned skull, which is competitive with the morphometrics based analysis.

Table 7.3: Performance comparison with Traditional Morphometrics

Estimation Method	Cranial Representation	Prediction Model	Prediction rate (%)
Morphometrics	22 estimation parameters	DFA and regression analysis	78.2 - 86.2
Proposed	**MeshLBP**	**KPCA-SVM**	**86**

7.3.2 Statistical Analysis

This method of analysis is another approach which has been used for sex determination from human skull, where principal component analysis (PCA) is applied for feature extraction and linear discriminant analysis (LDA) is used for classification. In this approach, the analysis can be based on the entire skull or on a few anatomical landmarks which have been manually annotated by forensic experts. The PCA/LDA based approach has been used by Luo et al. (2013a) for sex determination on dataset containing only Chinese samples. Since, the samples used by Luo et al. (2013a) is not publicly available, the PCA/LDA technique is applied on the dataset used in this thesis. In order to achieve this, it is important to establish a one-to-one correspondence between the sample skulls. As a result, skull registration is performed using non-rigid ICP. However, considering the size of the vertex points of each skull which affects the possibility of applying non-rigid ICP, all samples are first downsampled to 70000 faces. Then a sample skull is chosen as the model skull used for registering the remaining samples. Once the one-to-one correspondence has been establish, the vertex points are concatenated and place in a matrix.

Then 60 samples are drawn into the training set and 40 into the testing set. PCA is applied on the training samples to derive the eigenvectors and eigenvalues, which are then used to obtain the sub-space representations of the training and testing set. Once this has been achieved, linear discriminant analysis (LDA) is performed on the transformed training set, the derived parameters (weights) of LDA are used to assign the testing the samples into different classes. The results obtained are shown in Table 7.4. The results indicate that the proposed CV based

solution is better than PCA/LDA based technique, because the 3D shape representation is able to offer comprehensive description of the skull local details, unlike PCA which only represents the skull holistically. Thus, 3D shape features are more invariant than PCA-based features.

Table 7.4: Performance comparison with Traditional Morphometrics

Estimation Method	Cranial Representation	Prediction Model	Prediction rate (%)
Statistical Analysis Luo et al. (2013a)	PCA	LDA	75
Proposed	**MeshLBP**	**KPCA-SVM**	**86**

7.3.3 Patch Representation and Keypoints Analysis

Another analysis performed in this thesis is the integration of patch representation and automatically detected keypoints. The keypoints in this case are calculated as the centroid of each local patch. As such, as the skull is partitioned into different local regions, the centre local keypoint of that particular patch is computed and both features are fused for sex prediction. The essence of this approach is to provide a solution which mimics a hybrid representation of morphological and morphometric analysis. To achieve this, the MeshLBP local descriptor is computed for each patch and the centroid of the patch is considered as the keypoint. MeshLBP is considered for this analysis because the other local descriptor (spin image and LD-SIFT) are already keypoint based. Once the features have been extracted and fused with the keypoints, KPCA is then applied for dimensionality reduction and SVM for classification. The result attained is illustrated in Figure 7.1.

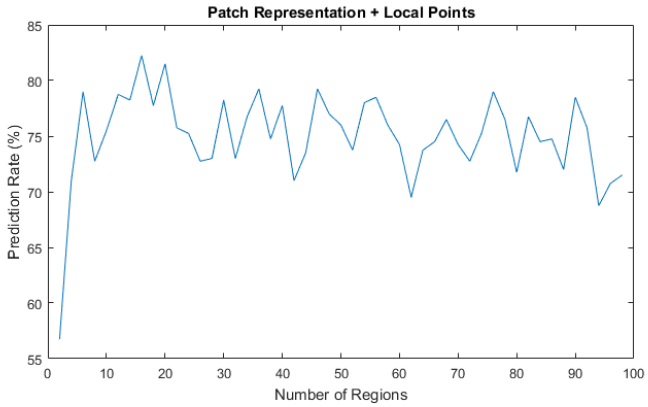


Figure 7.1: Figure of prediction rate for each partition.

The best prediction result using this approach is 83% obtained with 17 skull regions. It is can be noticed from the results that the local keypoints do not provide any improvement on the performance of the MeshLBP local descriptor, because

the information provided by the each keypoint is inconsequential on the shape representation derived from the patches.

Chapter 8

Discussion and Conclusion

> "There is no real ending. Its just the place where you stop the story."
>
> Frank Herbert

Determination of sex is very crucial in forensic examination of skeletal remains. It has received vast amount of attention over decades. Traditionally, the pelvic region is a body part which has been primarily considered for such analysis. However, due to degradation of the skeletal remains after preservation of the subjects over a period time, the skull region has been suggested as the second best predictor of sex based on comprehensive research findings. Conventional methods for sex determination are based on the study of both morphological and morphometric variations of the skull. These methods visually assess specific anatomical regions of the skull or make use of a few set of landmarks that are carefully localized on the skull. Eventually both approaches consider DFA as a model to assign the skulls into male and females classes.

This thesis diverged from the conventional approach and investigated how modern feature learning techniques, which have gained prominence in the machine learning and computer vision community, can possibly be utilized to extract more comprehensive features for predicting sex from human skull data. In order to achieve this, the research work has been divided into different aspects focusing on different objectives.

To begin with, a set of experiments is performed to assess the influence of different methods of skull partitioning on sex estimation results. The skulls are partitioned into local regions since this is a common practice by forensic anthropologists when assessing sex from human skull. As a result, 2 different partitioning methods have been introduced in this study: (1) Heursitic partition, which is simply division of the skull into geometric parts along the X,Y, and Z axes similar to the axial, coronal, and saggital axes. (2) Pseudo-anatomical partition which fairly resemble

the anatomical regions in standard forensic analysis. The evaluation is centralized around assessing different combination of 3D local shape descriptors, feature learning, and classification algorithms. The best results obtained is 86% for heuristic based partitioning using combination of MeshLBP+KPCA+SVM and 86.5% for pseudo-anatomical partitioning using combination of LD-SIFT+KPCA+SVM.

Furthermore, this thesis conducted an examination of how multiple information extracted from the skull can be induced to offer a comprehensive representation. To achieve this, the multiple kernel learning (MKL) is used to integrate different 3D local shape descriptors and the result obtained is inline with the previous results, with the best result of 86%.

To proceed further, some experiments are conducted to assess the importance of skull regions on estimation results. This is performed to advance conventional forensic methods, which literally utilize a stepwise combination of different skull sites for prediction. In this thesis, a 2 novel approaches (1) Novel Group Regression Model for Manifold Data and (2) Novel Group Regression Model for Manifold Data with Split Augmented Lagrangian Algorithm, which are based on the concept of group variable selection are proposed to automatically discover the important skull sites. The best result obtained under this approach is 84.5%.

From the results in this thesis, it can be concluded that the proposed approaches are competitive with standard forensic methods, where the prediction rate between 74% - 94% has generally been reported by forensic anthropologists (Garvin et al., 2014; Walker, 2008).

From the experiments conducted, it can be further suggested that multi-region representation should be considered for feature extraction instead of full representation of the skull data. However, careful selection of 3D local shape descriptors is crucial because not all descriptors performed effectively, for instance SIHKS, region area, and ShapeDNA did not perform accurately for sex determination. This is due to the fact that difference or variation between the skull of 2 different categories is very subtle. Even forensic experts do have high inter-observer error in some cases when using morphological assessment (Lewis and Garvin, 2016; Stevenson et al., 2009; Walrath et al., 2004). Despite the fact that the explored feature extraction methods exploit the topological properties shared by neighbouring vertex points in a 2D manifold (surface), depending on the representation from the entire skull, where each vertex is described independently would over complicate the description of the skull. Therefore, it is preferable to extract the features from different sub-regions. More importantly, feature representation of the entire vertex points should be summarized into small compact representation that capture both

local and global shape properties. In addition, it can be noticed that it is worth extracting information from multiple sources when available.

Finally learning based skull region selection should be considered ahead of stepwise selection, since learning based methods have the capability to scale to large size of data without demanding too much computational rigour, as opposed to stepwise method which will demand a large computational time when performing the sequential step-by-step combination of variables.

8.1 Future Works

In the thesis, several experiments have been conducted examining the problem of sex determination from human skull based on 3 main methods. In future works:

- Other methods which can be directly used to learn information from the original 3D points of the skull should be considered. This approaches could take the form of unsupervised or semi-supervised learning. The advantage of this is to ensure that the discovered features are primarily guided by the distribution of the data itself.

- Automatic approach to detect anatomical landmarks can be studied to assist in morphometric analysis. Few literature have reported the possibility of detecting anatomical landmarks automatically. However, these approaches have not attained the level of efficiency and robustness needed for a standalone forensic identification system.

- Methods to automatically derive semantic representations from the skull can be studied to aid morphological assessment without the need of human observers. It is well known that subjectivity has a significant effect on semantic terms derived by forensic observers. As such, it will be interesting to investigate how these descriptions can be discern automatically with no human influence.

- Further experiments using heuristic and pseudo-anatomical partitioning (including region selection) should be conducted for age and ancestry estimation.

- This research has shown that multiple information (features) from the skull can be fused from different sources. However, researchers should investigate how multiple skeletal regions can be combined to improve prediction performance. This is essential in cases where a particular skeletal region is incomplete, which could be compensated for through combination of another

skeletal regions. This aspect of research is still in its infancy and a recent report demonstrated the combination of skull and pelvic for sexual dimorphism (Best et al., 2017).

Appendix A

Prototype

The system prototype design for this research is depicted in Figure A.1. It consists of 5 main stages with several sub functions implementing each stage of the sex estimation system. The functionality of each button is described as follows:

- Select Data - loads a set of 2D CT slices and accumulates them in a 3D volumetric data. On the right panel of Figure A.1 is the 3D visualization of the coronal, axial, and saggital planes.
- Preprocess Data - is the button for removing noise, background objects, reconstruction of the volumetric data to 3D mesh, and mesh registration. An examples of a reconstructed skull after noise removal and mesh registration is shown in Figure A.2.
- Extract Mesh - computes the 3D shape primitive functions such mean curvature, Gaussian curvature, principal curvature, and shape index, as shown in Figure A.3. These functions are used in the next stage for extracting 3D shape descriptors.
- Extract features - compute different type of features such LD-SIFT, Spin Image, and MeshLBP as illustrated in Figure A.4.
- Learn/Classify - has two functions: (1) it learn compact and discriminative representation from the extracted features using KPCA or LPP. (2) it classifies the learned patterns into their respective demographic class such as sex, age, and ancestry as shown in the button panel of Figure A.4.

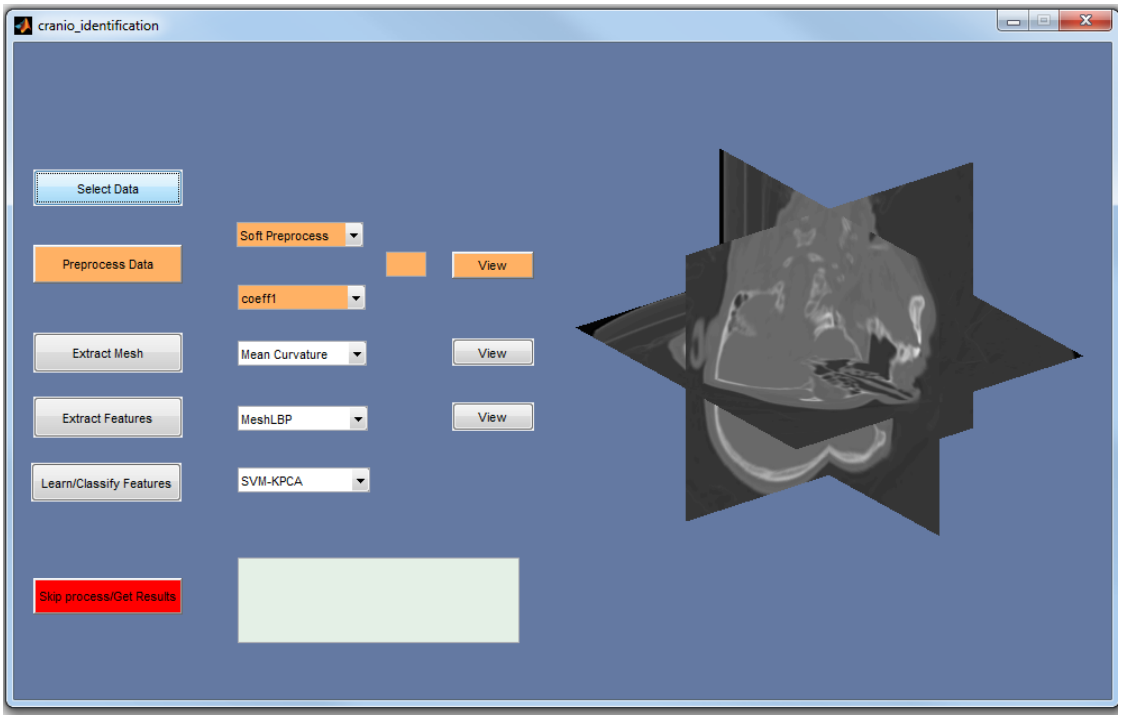


Figure A.1: Screenshot of the designed prototype showing the 3D planes of 3D CT volume

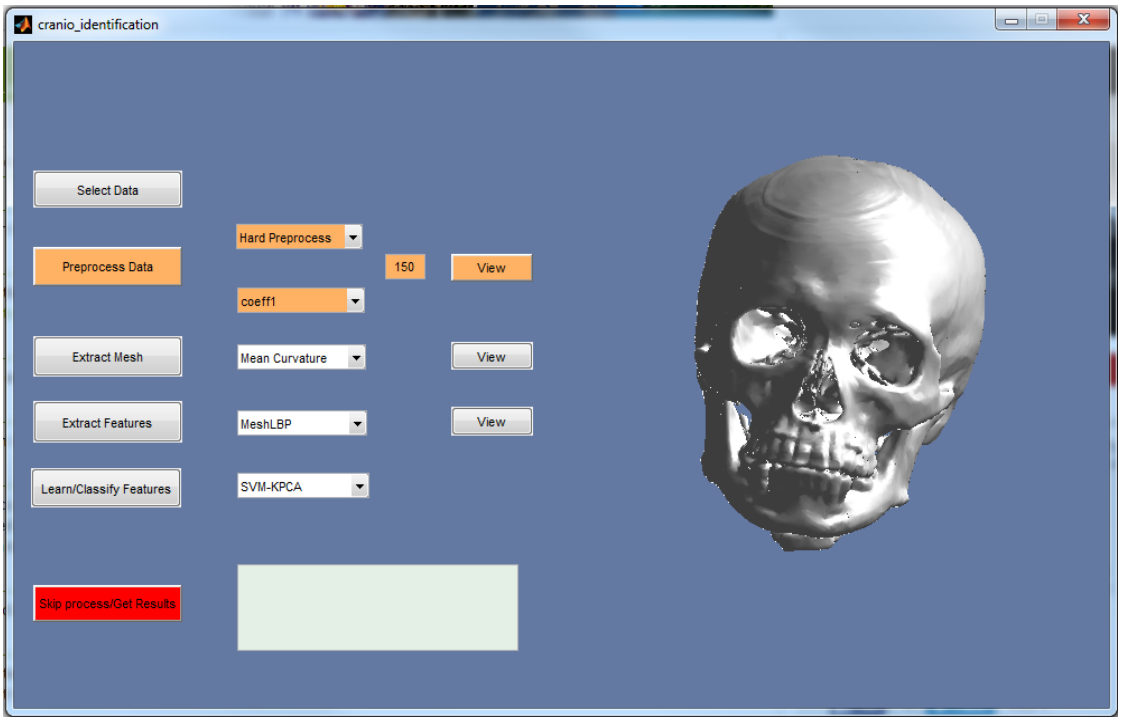


Figure A.2: Example of skull reconstruction

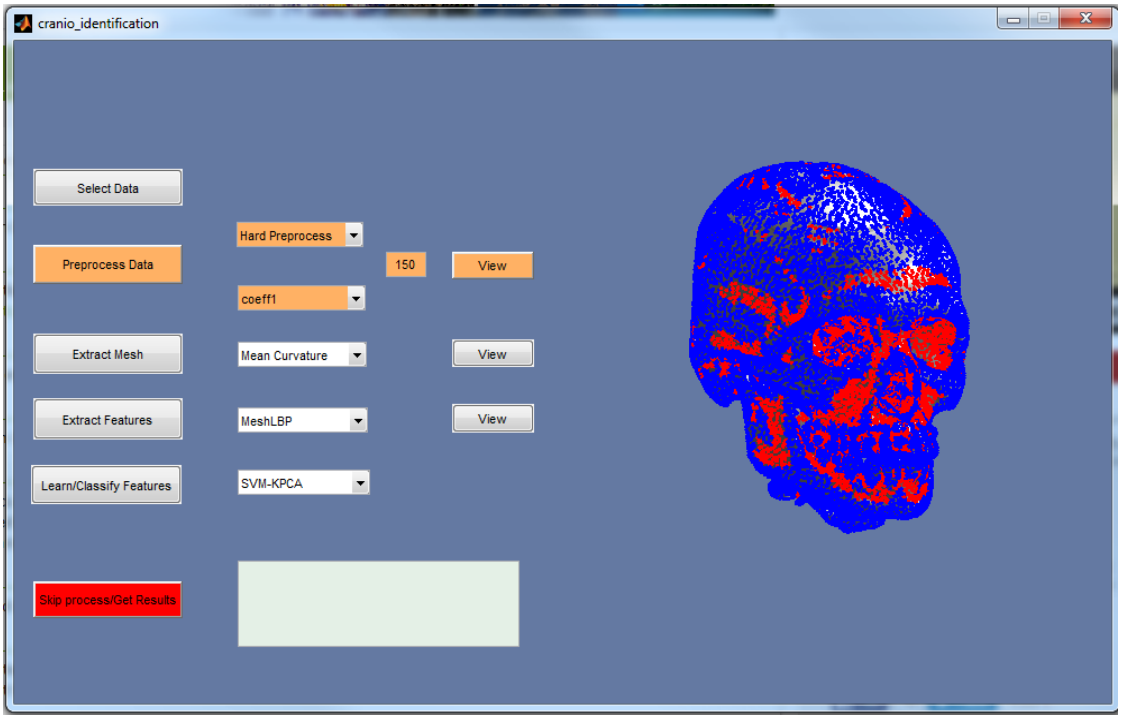


Figure A.3: Illustration of shape primitive function with an example depicting mean curvature

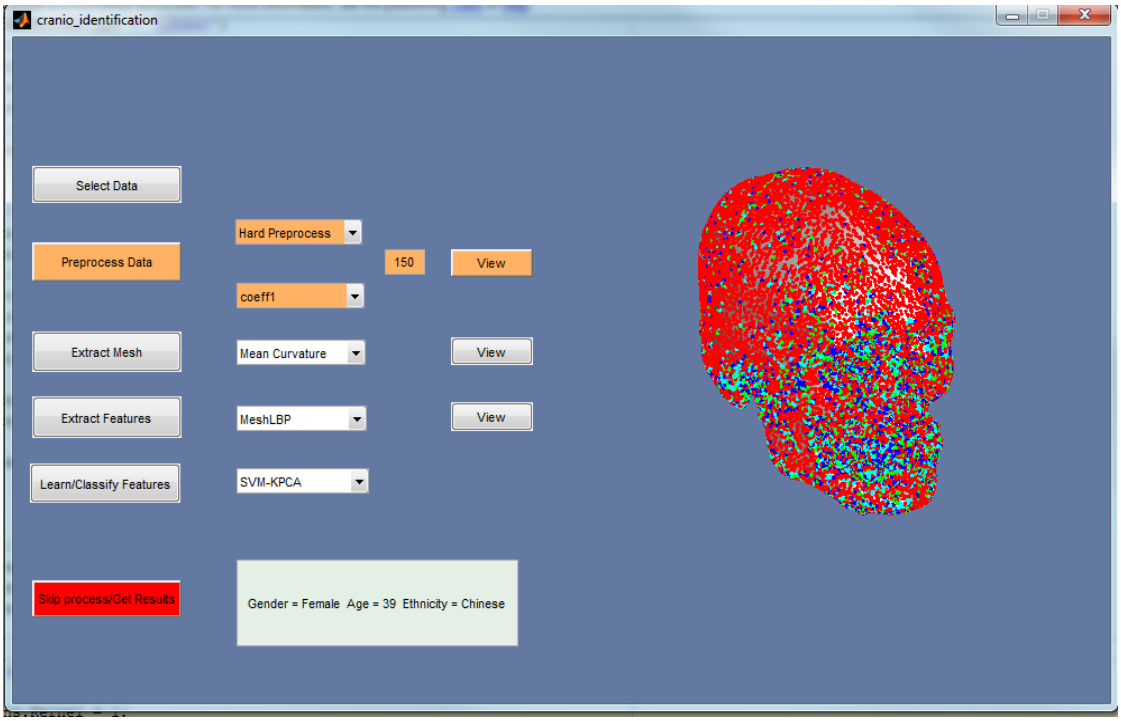


Figure A.4: shows the extracted MeshLBP and the prediction results

Bibliography

Abbas, H., Hicks, Y., Marshall, D., Zhurov, A. I., and Richmond, S. A. (2018). 3D morphometric perspective for facial gender analysis and classification using geodesic path curvature features. *Computational Visual Media*, 1–16.

Abdel Fatah, E. E., Shirley, N. R., Jantz, R. L., and Mahfouz, M. R. (2013). Improving sex estimation from crania using a novel threedimensional quantitative method. *J. Forensic Sci. Int.*, 59(3): 590-600.

Acsádi, G. Y. and Nemeskéri, J. (1970). *History of Human Life Span and Mortality.*

Adams, D. C., Rohlf, F. J., and Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the revolution'. *Italian Journal of Zoology*, 71(1):5–16.

Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. (2011). An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems. *IEEE Transactions on Image Processing*, 20(3):681–695.

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322.

Albanese, J. and Saunders, S. R. (2006). Is it possible to escape racial typology in forensic identification? *Forensic Anthropology and Medicine: Complementary Sciences From Recovery to Cause of Death*, pages 281–316.

Alexandre, L. A. (2010). Gender recognition: A multiscale decision fusion approach. *Pattern Recognition Letters*, 31(11):1422–1427.

Alley, T. R. (1988). *Social and applied aspects of face perception: An introduction.*

Anderson, M. F., Anderson, D. T., and Wescott, D. J. (2010). Estimation of adult skeletal age-at-death using the sugeno fuzzy integral. *American Journal of Physical Anthropology*, 142(1):30–41.

Ansary, T. F., Daoudi, M., and Vandeborre, J.-P. (2005). 3D Model Retrieval Based on Adaptive Views Clustering. In *Pattern Recognition and Image Analysis*, pages 473–483.

Arigbabu, O. A., Ahmad, S. M. S., Adnan, W. A. W., and Yussof, S. (2015). Recent advances in facial soft biometrics. *The Visual Computer*, 31(5):513–525.

Arigbabu, O. A., Ahmad, S. M. S., Adnan, W. A. W., Yussof, S., Iranmanesh, V., and Malallah, F. L. (2014). Gender recognition on real world faces based on shape representation and neural network. In *Proceedings of 2014 IEEE International Conference on Computer and Information Sciences (ICCOINS)*, pages 1–5.

Attene, M., Falcidieno, B., Spagnuolo, M., 2006a. *Hierarchical mesh segmentation based on fitting primitives. Visual Computer*, 22(3), 181–193.

Attene, M., Patané, G., Katz, S., Spagnuolo, M., Mortara, M., Tal, A. (2006). Mesh segmentation - A comparative study. *Proceedings - IEEE International Conference on Shape Modeling and Applications 2006, 7.*

Atmosukarto, I., Wilamowska, K., Heike, C., and Shapiro, L. G. (2010). 3D object classification using salient point patterns with application to craniofacial research. *Pattern Recognition*, 43(4):1502–1517.

Aulsebrook, W. A., Ican, M. Y., Slabbert, J. H., and Becker, P. (1995). Superimposition and reconstruction in forensic facial identification: A survey. *Forensic Science International*, 75(2-3):101–120.

Bakin, S. (1999). Adaptive regression and model selection in data mining problems.

Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research*, 9:1179-1225.

Ballihi, L., Amor, B. B., Daoudi, M., Srivastava, A., and Aboutajdine, D. (2012). Boosting 3-D-geometric features for efficient face recognition and gender classification. *IEEE Transactions on Information Forensics and Security*, 7(6):1766–1779.

Bekios-Calfa, J., Buenaposada, J. M., and Baumela, L. (2014). Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recognition Letters*, 36:228–234.

Berbar, M. A. (2013). Three robust features extraction approaches for facial gender classification. *The Visual Computer*, 30(1):19–31.

Besl, P. and McKay, N. (1992). A Method for Registration of 3-D Shapes.

Best, K. C., Garvin, H. M., Cabo, L (2017). An Investigation into the Relationship between Human Cranial and Pelvic Sexual Dimorphism. *Journal of Forensic Sciences*, 1–11.

Bidmos, M., Gibbon, V., and Štrkalj, G. (2010). Recent advances in sex identification of human skeletal remains in South Africa. *South African Journal of Science*, 106(11/12):1–7.

Bigoni, L., Velemínská, J., and Bržek, J. (2010). Three-dimensional geometric morphometric analysis of cranio-facial sexual dimorphism in a Central European sample of known sex. *HOMO- Journal of Comparative Human Biology*, 61(1):16–32.

Birkby, W. H., Fenton, T. W., and Anderson, B. E. (2008). Identifying Southwest Hispanics using nonmetric traits and the cultural profile. *Journal of forensic sciences*, 53(1):29–33.

Bookstein, F. L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical image analysis*, 1(3):225–243.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers *Foundations and Trends in Machine Learning*, 3(1): 1–122.

Boyd, K. L., Villa, C., and Lynnerup, N. (2015). The Use of CT Scans in Estimating Age at Death by Examining the Extent of Ectocranial Suture Closure. *Journal of Forensic Sciences*, 60(2):363–369.

Broca, P. (1875). Instructions craniologiques et craniométriques de la Société d'anthropologie de Paris. *Bulletins de la Société d'anthropologie de Paris*, 10:337–367.

Bronstein, A. M., Bronstein, M. M., and Ovsjanikov, M. (2010). 3D features, surface descriptors, and object descriptors. In *3D Imaging, Analysis, and Applications*, pages 1–27.

Brues, A. M. (1990). The once and future diagnosis of race. In *Skeletal Attribution of Race Methods for Forensic Anthropology Maxwell Museum of Anthropology: University of New Mexico*, pages 1–9.

Brunelli, R. and Poggio, T. (1992). HyberBF Networks for Gender Classification. In *DARPA Image Understanding Workshop*, number 1140, pages 311–314.

Buikstra, J. E. and Ubelaker., D. H. (1994). Standards for data collection from human skeletal remains. In *Proceedings of a seminar at the Field Museum of Natural History (Arkansas Archaeology Research Series 44)*.

Burton, A. M., Bruce, V., and Dench., N. (1993). "What's the difference between men and women? Evidence from facial measurement.". *PERCEPTION-LONDON*, (22):153–153.

Cattaneo, C. (2007). Forensic anthropology: developments of a classical discipline in the new millennium. *Forensic Science International*, 165(2-3):185–193.

Casado, A. (2017). Quantifying Sexual Dimorphism in the Human Cranium: A Preliminary Analysis of a Novel Method. *Journal of Forensic Sciences*, 165(2-3):185–193.

Chen, D.-Y., Tian, X.-P., Shen, Y.-T., and Ouhyoung, M. (2003). On Visual Similarity Based 3D Model Retrieval. *Eurographics*, 22(3):223–232.

Chen, J., Shan, S., He, C., Zhao, G., Pietikäinen, M., Member, S., Chen, X., and Gao, W. (2009). WLD : A Robust Local Image Descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1705–1720.

Chen, X., Yuan, X., Yan, S., Tang, J., Rui, Y., and Chua, T. S (2011). Towards multi-semantic image annotation with graph regularized exclusive group lasso. *Proc. of 19th ACM int. conf. on Multimedia - MM '11*, 263–272.

Cheng, H. and Chung, S. M. (2016). Orthogonal moment-based descriptors for pose shape query on 3D point cloud patches. *Pattern Recognition*, 52:397–409.

Chiba, F., Makino, Y., Motomura, A., Inokuchi, G., Torimitsu, S., Ishii, N., Sakuma, A., Nagasawa, S., Saitoh, H., Yajima, D., Hayakawa, M., Odo, Y., Suzuki, Y., and Iwase, H. (2013). Age estimation by multidetector CT images of the sagittal suture. *International Journal of Legal Medicine*, 127(5):1005–1011.

Cignoni, P., Montani, C., and Scopigno, R. (1998). A comparison of mesh simplification algorithms. *Computers & Graphics*, 22(1):37–54.

Comas, J. (1960). *Manual of Physical Anthropology*.

Cootes, T. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 297:273–297.

Dalal, N. (2006). *Finding people in images and videos*. PhD thesis.

Darom, T. and Keller, Y. (2012). Scale Invariant Features for 3D Mesh Models. *IEEE Transactions on Image Processing*, 21(5):1–32.

De Angelis, D., Gaudio, D., Guercini, N., Cipriani, F., Gibelli, D., Caputi, S., and Cattaneo, C. (2015). Age estimation from canine volumes. *La radiologia medica*, 120(8):731–736.

Demirkus, M. and Toews, M. (2010). Gender classification from unconstrained video sequences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 55–62.

Dhamecha, T. and Sankaran, A. (2011). Is gender classification across ethnicity feasible using discriminant functions? In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1 – 7.

Do, T.-N. and Fekete, J.-D. (2007). Large Scale Classification with Support Vector Machine Algorithms. In *Proceedings - 6th International Conference on Machine Learning and Applications, ICMLA 2007*, pages 148–153.

Dong, H., Deng, M., Wang, W., Zhang, J., Mu, J., and Zhu, G. (2015). Sexual dimorphism of the mandible in a contemporary Chinese Han population. *Forensic Science International*, 255:9–15.

Tran, D., Lubomir, B., Fergus, R., Torresani, L., and Paluri, M. (2015). Learning Spatiotemporal Features with 3D Convolutional Networks. *IEEE CVPR*, 4489–4497.

Dwight, T. (1890). The Closure of the Cranial Sutures as a Sign of Age. *The Boston Medical and Surgical Journal*, 122(17):389–392.

Farkas, L. G. (1994). *Anthropometry of the Head and Face*.

Fellous, J. M. (1997). Gender discrimination and prediction on the basis of facial metric information. *Vision Research*, 37(14):1961–1973.

Franklin, D. (2010). Forensic age estimation in human skeletal remains: Current concepts and future directions. *Legal Medicine*, 12(1):1–7.

Franklin, D., Cardini, A., Flavel, A., and Kuliukas, A. (2012). The application of traditional and geometric morphometric analyses for forensic quantification of sexual dimorphism: Preliminary investigations in a Western Australian population. *International Journal of Legal Medicine*, 126(4):549–558.

Franklin, D., Cardini, A., Flavel, A., and Kuliukas, A. (2013). Estimation of sex from cranial measurements in a Western Australian population. *Forensic Science International*, 229(1-3):158.e1–158.e8.

Franklin, D., Freedman, L., and Milne, N. (2005). Sexual dimorphism and discriminant function sexing in indigenous South African crania. *HOMO - Journal of Comparative Human Biology*, 55(3):213–228.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). A note on the group lasso and a sparse group lasso. *arXiv:1001.0736 [math, stat]*, page 8.

Fu, S., He, H., and Hou, Z.-G. (2014). Learning Race from Face: A Survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(12):2483–2509.

Fu, Y., Guo, G., and Huang, T. S. (2010). Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976.

Fu, Y. and Huang, T. S. (2008). Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, 10(4):578–584.

Galdames, I. C. S., Matamala, D. A. Z., Smith, R. L., SUAZO, G., ZAVANDO, M., and SMITH, R. (2008). Sex determination using mastoid process measurements in Brazilian skulls. *Int J Morphol*, 26(4):941–4.

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis*, 54(4):1167–1178.

Garvin, H. M. and Passalacqua, N. V. (2012). Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. *Journal of Forensic Sciences*, 57(2):427–433.

Garvin, H. M., Sholts, S. B., and Mosca, L. A. (2014). Sexual dimorphism in human cranial trait scores: Effects of population, age, and body size. *American Journal of Physical Anthropology*, 154:259–269.

Garvin, H. M., and Klales, A. R. (2018). A Validation Study of the Langley et al. (2017) Decision Tree Model for Sex Estimation. *J. Forensic Sci.*, 63(4):1243–1251.

Gehler, P. and Nowozin, S. (2009). On feature combination for multiclass object classification. *Computer Vision, 2009 IEEE 12th International Conference on*, (Iccv):221–228.

Geng, X., Zhou, Z. H., and Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2234–2240.

Giles, E. (1964). Sex determination by discriminant function analysis of the mandible. *American Journal of Physical Anthropology*, 22(2):129–135.

Giles, E. and Elliot, O. (1962). Race identification from cranial measurements. *Journal of Forensic Sciences*, 7(2):147–167.

Giles, E. and Elliot, O. (1963). Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology*, 21(1):53–68.

Gill, G. W. and Gilbert, B. M. (1990). Race identification from the midfacial skeleton: American Blacks and Whites. *Skeletal attribution of race: Methods for forensic anthropology. Anthropological Papers of the Maxwell Museum of Anthropology 4*, pages 47–53.

Giurazza, F., Del Vescovo, R., Schena, E., Battisti, S., Cazzato, R. L., Grasso, F. R., Silvestri, S., Denaro, V., and Zobel, B. B. (2012). Determination of stature from skeletal and skull measurements by CT scan evaluation. *Forensic Science International*, 222(1-3):398.e1–398.e9.

Golomb, B., Lawrence, D., and Sejnowski, T. (1990). SEXNET: A Neural Network Identifies Sex From Human Faces. In *Proceedings of 1990 conference on Advances in neural information processing systems 3 (NIP-3). Richard P. Lippmann, John E. Moody, and David S. Touretzky (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.*, pages 572–577.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations.*

Gönen, M. and Alpaydn, E. (2011). Multiple Kernel Learning Algorithms. *Journal of Machine Learning Research*, 12:2211–2268.

Gonzalez, P. N., Bernal, V., and Perez, S. I. (2011). Analysis of sexual dimorphism of craniofacial traits using geometric morphometric techniques. *International Journal of Osteoarchaeology*, 21(1):82–91.

Gonzlez-Briones, A., Villarrubia, G., De Paz, J. F., and Corchado, J. M. (2018). A multi-agent system for the classification of gender and age from images. *Computer Vision and Image Understanding*, January, In Press.

Graw, M., Czarnetzki, A., and Haffner, H. T. (1999). The form of the supraorbital margin as a criterion in identification of sex from the skull: Investigations based on modern human skulls. *American Journal of Physical Anthropology*, 108:91–96.

Guo, G., Fu, Y., Dyer, C. R., and Huang, T. S. (2008). Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188.

Guyomarc'h, P. and Bruzek, J. (2011). Accuracy and reliability in sex determination from skulls: A comparison of Fordisc?? 3.0 and the discriminant function analysis. *Forensic Science International*, 208:30–35.

Hammond, P. (2007). The use of 3D face shape modelling in dysmorphology. *Archives of disease in childhood*, 92(12):1120–1126.

Han, H., Otto, C., Liu, X., and Jain, A. (2015). Demographic Estimation from Face Images: Human vs. Machine Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1148–1161.

Han, X., Ugail, H., and Palmer, I. (2009). Gender Classification Based on 3D Face Geometry Features Using SVM. In *2009 International Conference on CyberWorlds*, pages 114–118.

Harth, S., Obert, M., Ramsthaler, F., Reuss, C., Traupe, H., and Verhoff, M. (2010). Ossification degrees of cranial sutures determined with flat-panel computed tomography: narrowing the age estimate with extrema. *Journal of forensic sciences*, 55(3):690–694.

Harth, S., Obert, M., Ramsthaler, F., Reuss, C., Traupe, H., and Verhoff, M. a. (2009). Estimating age by assessing the ossification degree of cranial sutures with the aid of Flat-Panel-CT. *Legal medicine (Tokyo, Japan)*, 11 Suppl 1:S186–9.

He, X. and Niyogi, P. (2004). Locality preserving projections. In *Neural information processing systems*, volume 16, page 153.

Hefner, J. T. (2003). *Assessing Nonmetric Cranial Traits currently used in forensic determination of Ancestry*. PhD thesis.

Hefner, J. T. (2007). *The statistical determination of ancestry using cranial nonmetric traits*. PhD thesis.

Hefner, J. T. (2009). Cranial Nonmetric Variation and Estimating Ancestry. *Journal of Forensic Sciences*, 54(5):985–995.

Hefner, J. T., Emanovsky, P. D., Byrd, J., and Ousley, S. D. (2007). The value of experience, education, and methods in ancestry prediction. In *Proceedings of the 59th Annual Meeting of the American Academy of Forensic Sciences. American Academy of Forensic Sciences.*

Hefner, J. T. and Ousley, S. D. (2006). Morphoscopic traits and the statistical determination of ancestry II. In *Proceedings of the 58th Annual Meeting of the American Academy of Forensic Sciences.*, pages 20–25.

Hefner, J. T., Spradley, M. K., and Anderson, B. (2014). Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, 59(3):583–589.

Hershkovitz, I., Latimer, B., Dutour, O., Jellema, L. M., Wish-Baratz, S., Rothschild, C., and Rothschild, B. M. (1997). Why do we fail in aging the skull from the sagittal suture? *American Journal of Physical Anthropology*, 103(April 1996):393–399.

Hestenes, M. R. (1969). Multiplier and gradient methods *Journal of optimization theory and applications*, 4(5):303–320.

Hjelmås, E. and Low, B. K. (2001). Face Detection: A Survey. *ACM Computing Surveys*, 83(3):236–274.

Hrdlička, A. (1939). *Practical Anthropometry*.

Hu, Y., Yan, J., and Shi, P. (2010). A fusion-based method for 3D facial gender classification. In *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, volume 5, pages 369–372.

Huang, G.-B. (2014). An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. *Cognitive Computation*, 6(3):376–390.

Huang, J., Breheny, P., Ma, S. (2012). A Selective Review of Group Selection in High-Dimensional Models. *Statistical Science*, 27(4):481–499.

Huang, J., Breheny, P., and Ma S. (2012). A Selective Review of Group Selection in High-Dimensional Models. *Stat. Sci.*, 27(4): 481–499

Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501.

Huynh, T., Min, R., Dugelay, J. L. (2013). An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 133–145.

Ibáñez, O., Vicente, R., Navega, D., Wilkinson, C., Jayaprakash, P., Huete, M., Briers, T., Hardiman, R., Navarro, F., Ruiz, E., Cavalli, F., Imaizumi, K., Jankauskas, R., Veselovskaya, E., Abramov, A., Lestón, P., Molinero, F., Cardoso, J., Çadr, A., Humpire, D., Nakanishi, Y., Zeuner, A., Ross, A., Gaudio, D., and Damas, S. (2015). Study on the performance of different craniofacial superimposition approaches (II): Best practices proposal. *Forensic Science International*.

Ibrahim, A., Alias, A., Nor, F. M., Swarhib, M., Abu Bakar, S. N., Das, S., Buran, C., and Sayin, I. (2017). Study of sexual dimorphism of Malaysian crania: an important step in identification of the skeletal remains. *Anatomy & Cell Biology*, 50(2):86.

Ican, M. Y. and Loth., S. R. (1986). Determination of age from the sternal rib in white males: a test of the phase method. *Journal of Forensic Science*, 31(1):122–132.

Ilayperuma, I. (2010). On the Prediction of Personal Stature from Cranial Dimensions. *International Journal of Morphology*, 28(4):1135–1140.

Jagannathan, A., and Miller, E. L. (2007). Three-dimensional surface mesh segmentation using curvedness-based region growing approach. *IEEE Transactions on pattern analysis and machine intelligence*, 29(12), 2195–2204.

Jia, S., Landsall-Welfare, T., and Cristianini, N. (2016). Gender Classification by Deep Learning on Millions of Weakly Labelled Images. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 462–467.

Jiménez-Arenas, J. M. and Esquivel, J. A. (2013). Comparing two methods of univariate discriminant analysis for sex discrimination in an Iberian population. *Forensic science international*, 228:175.e1–175.e4.

Johnson, A. (1997). Spin-images: a representation for 3-D surface matching. *Image (Rochester, N.Y.)*, pages 1–7.

Jung, H., and Woo, E. J (2016). Evaluation of Mastoid Process as Sex Indicator in Modern White Americans using Geometric Morphometrics. *Journal of forensic and legal medicine*, 61(4):1029–1033.

Kanchan, T., Gupta, A., and Krishan, K. (2013). Estimation of sex from mastoid triangle - a craniometric analysis. *Journal of forensic and legal medicine*, 20(7):855–60.

Katz, S., and Tal, A. (2003). Hierarchical mesh decomposition using fuzzy clustering and cuts *ACM*, 22: 954–961.

Kazhdan, M. and Funkhouser, T. (2002). Harmonic 3D shape matching. *ACM SIGGRAPH 2002 conference abstracts ...*, 12:2001.

Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. (2003). Rotation invariant spherical harmonic representation of 3D shape descriptors. *Proceedings of the 2003 ...*, pages 156–165.

Khan, F. S., Van De Weijer, J., Anwer, R. M., Felsberg, M., and Gatta, C. (2014). Semantic pyramids for gender and action recognition. *IEEE Transactions on Image Processing*, 23(8):3633–3645.

Khan, S. A., Nazir, M., Akram, S., and Riaz, N. (2011). Gender classification using image processing techniques: A survey. In *Proceedings of the 2011 IEEE 14th International Multitopic Conference (INMIC)*, pages 25–30.

Konigsberg, L. W. and Hens, S. M. (1998). Use of ordinal categorical variables in skeletal assessment of sex from the cranium. *American journal of physical anthropology*, 107(1):97–112.

Konigsberg, L. W., Herrmann, N. P., Wescott, D. J., and Kimmerle, E. H. (2008). Estimation and evidence in forensic anthropology: Age-at-death. *Journal of Forensic Sciences*, 53(3):541–557.

Krogman, W. M. (1962). *The Human Skeleton in Forensic Medicine*. American Association for the Advancement of Science.

Krüger, G. C., L'Abbé, E. N., Stull, K. E., and Kenyhercz, M. W. (2015). Sexual dimorphism in cranial morphology among modern South Africans. *International Journal of Legal Medicine*, 129(4):869–875.

Kwon, Y. H. and Lobo, N. d. V. (1994). Age Classification from Facial Images. In *Proceedings, IEEE*, pages 762–768.

Kwon, Y. H. and Lobo, N. d. V. (1999). Age classification from facial images. *Computer Vision and Pattern Recognition*, 74(1):1–21.

Lanitis, A. (2002). On the significance of different facial parts for automatic age estimation. *2002 14th International Conference on Digital Signal Processing Proceedings. DSP 2002 (Cat. No.02TH8628)*, 2:27–30.

Lanitis, A., Taylor, C. J., and Cootes, T. F. (2002). Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455.

Jacob, L., Obozinski, G., and Vert, J. P. (2009). Group Lasso with Overlaps and Graph Lasso. *2009 ICML*, pages 433–440.

Lai, Y. K., Hu, S. M., Martin, R. R., and Rosin, P. L. (2008). Fast mesh segmentation using random walks. *In Proceedings of the 2008 ACM symposium on Solid and physical modeling*, pp. 183–191.

Lavagetto, F. and Pockaj, R. (2001). An efficient use of MPEG-4 FAP interpolation for facial animation at 70 bits/frame. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(10):1085–1097.

Langley, N. R., Dudzik, B., Cloutier, A. (2017). A Decision Tree for Nonmetric Sex Assessment from the Skull. *Journal of Forensic Sciences*, 61(3):743–751.

Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2169–2178.

Lewis, C. J. and Garvin, H. M. (2016). Reliability of the Walker Cranial Nonmetric Method and Implications for Sex Estimation. *Journal of Forensic Sciences*, 61(3):743–751.

Lo, L.-J. and Chen, Y.-R. (2003). Three-dimensional computed tomography imaging in craniofacial surgery: morphological study and clinical applications. *Chang Gung medical journal*, 26(1):1–11.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4):163–169.

Lowe, D. G. (2004). Distinctive Image Features from. *Distinctive Image Features from Scale-Invariant Keypoints*, 60(2):91–110.

Lu, X., Chen, H., and Jain, A. (2005). Multimodal facial gender and ethnicity identification. *Advances in Biometrics*, pages 554–561.

Luo, L., Chang, L., Liu, R., and Duan, F. (2013a). Morphological Investigations of Skulls for Sex Determination Based on Sparse Principal. In *Biometric Recognition, Lecture Notes in Computer Sciences*, pages 449–456.

Luo, L., Wang, M., Tian, Y., Duan, F., Wu, Z., Zhou, M., and Rozenholc, Y. (2013b). Automatic Sex Determination of Skulls Based on a Statistical Shape Model. *Computational and MathematicalMethods in Medicine*, 2013:1–7.

Liu, R., and Zhang, H. (2004). Segmentation of 3D meshes through spectral clustering. *In Proceedings of 12th IEEE Pacific Conference on Computer Graphics and Applications*, pages 298–305.

Zhang, H., and Liu, R. (2005). Mesh segmentation via recursive and visually salient spectral cuts. *In Proc. of vision, modeling, and visualization*, pp. 429–436.

Luu, K., Ricanek, K., Bui, T. D., and Suen, C. Y. (2009). Age estimation using Active Appearance Models and Support Vector Machine regression. In *Proc. IEEE Intl. Conf. on Biometrics: Theory, Applications, and Systems*, pages 1–5.

Majid, A., Khan, A., and Mirza, A. (2003). Gender classification using discrete cosine transformation: A comparison of different classifiers. In *7th International Multi Topic Conference*, pages 59–64.

Mansanet, J., Albiola, A., and Paredes, R. (2015). Local Deep Neural Networks for gender recognition. *Pattern Recognition Letters*, 70:80–86.

Mansoor Roomi, S. M., Virasundarii, S. L., Selvamegala, S., Jeevanandham, S., and Hariharasudhan, D. (2011). Race classification based on facial features. In *Proceedings - 2011 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2011*, pages 54–57.

McGraw, T., (2015). Sparse Non-negative Matrix Factorization for Mesh Segmentation. *International Journal of Image and Graphics*, 16(1), 1–16.

Mehta, M., Saini, V., Nath, S., Patel, M. N., and Menon, S. K. (2014). CT scan images to determine the origin from craniofacial indices for Gujarati population. *Journal of Forensic Radiology and Imaging*, 2(2):64–71.

Meindl, R. S. and Lovejoy, C. O. (1985). Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American journal of physical anthropology*, 68(1):57–66.

Meier, L., Van De Geer, S., Bhlmann, P. M. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 70(1):53–71.

Min L., Jian Z., Yulan G., and Yanxin M. (2016). Accelerated coherent point drift for automatic three-dimensional point cloud registration. *IEEE Geoscience and Remote Sensing Letters*, 13(2): 162–166.

Modesto, C.-S., Lorenzo-Navarro, J., and Ramón-Balmaseda, E. (2013). Improving Gender Classification Accuracy in the Wild. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 8259:270–277.

Murphy, R. E., and Garvin, H. M. (2017). A Morphometric Outline Analysis of Ancestry and Sex Differences in Cranial Shape. *Journal of forensic sciences*, 16–21.

Myronenko, A. and Song, X. (2010). Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell*, 32(12):2262–2275.

Musilov, B., Dupej, J., Velemnsk, J., Chaumoitre, K., Bruzek, J. (2017). Exocranial surfaces for sex assessment of the human cranium. *Forensic Science International*, 16–21.

Ni, B., Song, Z., and Yan, S. (2009). Web image mining towards universal age estimator. In *Proceedings of ACM international conference on Multimedia*, pages 85–94.

Ohbuchi, R., Minamitani, T., and Takei, T. (2005). Shape-similarity search of 3D models by using enhanced shape functions. *International Journal of Computer Applications in Technology*, 23(2/3/4):70.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.

Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. (2002). Shape distributions. *ACM Transactions on Graphics*, 21(4):807–832.

Paiva, L. A. S. and Segre, M. (2003). Sexing the human skull through the mastoid process. *Revista do Hospital das Clínicas*, 58(1):15–20.

Parsons, F. G. and Box, C. R. (1905). The relation of the cranial sutures to age. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 5(vol. 11):301–316.

Perizonius, W. (1984). Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (a.d. 18831909). *Journal of Human Evolution*, 13(2):201–216.

Pinto, S. C. D., Urbanová, P., and Cesar, R. M. (2016). Two-Dimensional Wavelet Analysis of Supraorbital Margins of the Human Skull for Characterizing Sexual Dimorphism. *IEEE Transactions on Information Forensics and Security*, 11(7):1542–1548.

Popa, F. M., tefnescu, C. L., and Corici, P. D. (2009). Forensic value of mandibular anthropometry in gender and age estimation. *Romanian Journal of Legal Medicine*, 17(1):45–50.

Powell, M. J (1969) A method for non-linear constraints in minimization problems. *UKAEA*

Qi, C. R., Su, H., Mo, K., and Guibas, L. J.(2016). PointNet : Deep Learning on Point Sets for 3D Classification and Segmentation Big Data + Deep Representation Learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 652–660.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *In Advances in Neural Information Processing Systems*, 5105–5114.

Ramanathan, N. and Chellappa, R. (2006). Modeling age progression in young faces. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:387–394.

Ramsthaler, F., Kettner, M., Gehl, A., and Verhoff, M. (2010). Digital forensic osteology: Morphological sexing of skeletal remains using volume-rendered cranial CT scans. *Forensic Science International*, 195(1-3):148–152.

Ramsthaler, F., Kreutz, K., and Verhoff, M. A. (2007). Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. *International Journal of Legal Medicine*, 121(6):447–482.

Ramamoorthy, B., Pai, M., Prabhu, L. V., Muralimanju, B. V., and Rai, R. (2015). Assessment of craniometric traits in South Indian dry skulls for sex determination.. *Journal of forensic and legal medicine*, 37:8–14.

Rao, N., Nowak, R., Cox, C., and Rogers, T. (2016). Classification with the sparse group lasso.. *IEEE Transactions on Signal Processing*, 64(2), 448–463.

Reuter, M., Wolter, F. E., and Peinecke, N. (2006). Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. *CAD Computer Aided Design*, 38(4):342–366.

Reuter, M. (2010). Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions. *International Journal of Computer Vision*, 89(2-3), 287–308.

Rhee, C.-H., Shin, S. M., Choi, Y.-S., Yamaguchi, T., Maki, K., Kim, Y.-I., Kim, S.-S., Park, S.-B., and Son, W.-S. (2015). Application of statistical shape analysis for the estimation of bone and forensic age using the shapes of the 2nd, 3rd, and 4th cervical vertebrae in a young Japanese population. *Forensic Science International*.

Rhine, S. (1990). Non-metric skull racing. *Skeletal attribution of race*, pages 9–20.

Rodrguez, P., Cucurull, G., Gonfaus, J. M., Roca, F. X., and Gonzalez, J. (2017). Age and gender recognition in the wild with deep attention. *Pattern Recognition*, 72:563–571.

Rogers, T. (1991). *Sex determination and age estimation: skeletal evidence from St. Thomas' Cemetry*. PhD thesis.

Rosas, A. and Bastir, M. (2002). Thin-plate spline analysis of allometry and sexual dimorphism in the human craniofacial complex. *American Journal of Physical Anthropology*, 117(3):236–245.

Salah, S. H. and Du, H. (2013). Fusing Local Binary Patterns with Wavelet Features for Ethnicity Identification. *International Journal of Computer, Information Science and Engineering*, 7(7):424–430.

Santos, F., Guyomarc'h, P., and Bruzek, J. (2014). Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines. *Forensic Science International*, 245:204.e1–204.e8.

Sauer, N. J. (1992). Forensic anthropology and the concept of race: If races don't exist, why are forensic anthropologists so good at identifying them? *Social Science and Medicine*, 34:107–111.

Scheuer, L. (2002). Application of osteology to forensic medicine. *Clinical Anatomy*, 15(4):297–312.

Schmitt, A., Cunha, E., and Pinheiro, J. (2006). *Forensic Anthropology and Medicine*.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319.

Scovanner, P., Saad A., and Mubarak S. (2009). A 3-dimensional sift descriptor and its application to action recognition. *Proceedings of the 15th ACM international conference on Multimedia*, pages 357–360.

Selesnick, I. (2011). L1-norm penalized least squares with SALSA. *Connexions*.

Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437.

Shan, C., Gong, S., and McOwan, P. W. (2008). Fusing gait and face cues for human gender recognition. *Neurocomputing*, 71(10-12):1931–1938.

Shapira, L., Shamir, A., and Cohen-Or, D. (2008). Consistent mesh partitioning and skeletonisation using the shape diameter function. *The Visual Computer*, 24(4), 249.

Shapiro, L. G., Wilamowska, K., Atmosukarto, I., Wu, J., Heike, C., Speltz, M., and Cunningham, M. (2009). Shape-based classification of 3D head data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5716 LNCS:692–700.

Shamir, A. 2008. A survey on mesh segmentation techniques. *Computer Graphics Forum*, 27(6), 1539–1556.

Sharma, A., Horaud, R. P., Knossow, D., and Von Lavante, E. (2009). Mesh segmentation using Laplacian eigenvectors and gaussian mixtures. *In AAAI Fall Symposium: Manifold Learning and Its Applications*.

Smeets, D., Keustermans, J., Vandermeulen, D., and Suetens, P. (2013). MeshSIFT: Local surface features for 3D face recognition under expression variations and partial data. *Computer Vision and Image Understanding*, 117(2):158–169.

Spradley, M. K. and Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, 56(2):289–296.

Stevenson, J. C., Mahoney, E. R., Walker, P. L., and Everson, P. M. (2009). Technical note: prediction of sex based on five skull traits using decision analysis (CHAID). *American journal of physical anthropology*, 139(3):434–41.

Stewart, T. D. (1979). *Essentials of forensic anthropology*.

Steyn, M. and Ican, M. Y. (1998). Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Science International*, 98(1-2):9–16.

Stull, K. E., Kenyhercz, M. W., and L'Abbé, E. N. (2014). Ancestry estimation in South Africa using craniometrics and geometric morphometrics. *Forensic Science International*, 245:206.e1–206.e7.

Sun, J., Ovsjanikov, M., and Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. *Eurographics Symposium on Geometry Processing*, 28(5):1383–1392.

Sun, Y., Zhang, M., Sun, Z., and Tan, T. (2018). Demographic analysis from biometric data: Achievements, challenges, and new frontiers. *IEEE transactions on pattern analysis and machine intelligence*, 40(2):332–351.

Tamura, S., Hideo, K., and Mitsumoto, H. (1996). Male / Female Identification from 8 x 6 Very Low Resolution Face Images by Neural Network. *Pattern Recognition*, 29(2):331–335.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B*, 58(1):267-288.

Tokola, R., Mikkilineni, A., and Boehnen, C. (2015). 3D face analysis for demographic biometrics. In *IEEE International Conference on Biometrics (ICB)*.

Tsuchie, S., Hosino, T., and Higashi, M. (2014). High-quality vertex clustering for surface mesh segmentation using Student-t mixture model. *Computer-Aided Design*, 46, 69–78.

Ueki, K., Hayashida, T., and Kobayashi, T. (2006). Subspace-based age-group classification using facial images under various lighting conditions. *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006:43–48.

Ullah, I., Hussain, M., Muhammad, G., Aboalsamh, H., Bebis, G., and Mirza, A. M. (2012). Gender Recognition from Face Image with Local WLD Descriptor. In *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 11–13.

Vieira, M., and Shimada, K. (2005). Surface mesh segmentation and smooth surface extraction through region growing. *Computer aided geometric design*, 22(8), 771–792.

Viola, P. and Jones, M. J. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137–154.

Vo, A. V., Truong-Hong, L., Laefer, D. F., and Bertolotto, M. (2015). Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 88–100.

Vranic, D. (2003). An improvement of rotation invariant 3D-shape based on functions on concentric spheres. *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, 3:757–760.

Vranic, D. and Saupe, D. (2002). Description of 3D-shape using a complex function on the sphere. *Proceedings - 2002 IEEE International Conference on Multimedia and Expo, ICME 2002*, 1:177–180.

Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, 136(1):39–50.

Walrath, D. E., Turner, P., and Bruzek, J. (2004). Reliability test of the visual assessment of cranial traits for sex determination. *American Journal of Physical Anthropology*, 125(2):132–137.

Wang, J., Li, J., Lee, C., and Yau, W. (2010). Dense SIFT and Gabor descriptors-based face representation with applications to gender recognition. In *Int. Conf. Control, Automation, Robotics and Vision*, pages 7–10.

Wang, M. and Shizhong, L. (2009). Group Structure Influence on Group Lasso Consistency. In *IEEE Fifth International Conference on Natural Computation*, 6: 493–497.

Wang, Y. and Yuan, B. (2001). A novel approach for human face detection from color images under complex background. *Pattern Recognition*, 34(10):1983–1992.

Werghi, N., Berretti, S., and del Bimbo, A. (2015a). The Mesh-LBP : A Framework for Extracting Local Binary Patterns From Discrete Manifolds. *IEEE Transactions on Image Processing*, 24(1):220–235.

Werghi, N., Rahayem, M., and Kjellander, J. (2012). An ordered topological representation of 3D triangular mesh facial surface : concept and applications. *EURASIP Journal on Advances in Signal Processing*, 2012(144):1–20.

Werghi, N., Tortorici, C., Berretti, S., and Bimbo, A. D. (2015b). Computing Local Binary Patterns on Mesh Manifolds for 3D Texture Retrieval. *Cvpr*, 1:1–4.

Williams, B. A. and Rogers, T. L. (2006). Evaluating the Accuracy and Precision of Cranial Morphological Traits for Sex Determination. *Journal of Forensic Sciences*, 51(4):729–735.

Wolff, K., Vas, Z., Sótonyi, P., and Magyar, L. G. (2012). Skeletal age estimation in Hungarian population of known age and sex. *Forensic Science International*, 223(1-3).

Wohlhart, P. and Lepetit, V. (2015). Learning Descriptors for Object Recognition and 3D Pose Estimation. *IEEE CVPR*, 1:3109–3118.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227.

Wu, J., Zhang, C., Xue, T., Freeman, T. W., and Tenenbaum, J. B. (2016). Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *NIPS*.

Xia, B., Amor, B. B., Drira, H., Daoudi, M., Ballihi, L. (2013). Gender and 3D Facial Symmetry : What's the Relationship ? *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition*.

Xia, B., Amor, B. B., Daoudi, M., and Drira, H. (2014). Can 3D Shape of the Face Reveal your Age ? *International Conference on Computer Vision Theory and Applications*.

Xia, B., Amor, B. B., Drira, H., Daoudi, M., Ballihi, L. (2015). Combining face averageness and symmetry for 3D-based gender classification *Pattern Recognition*, 48(3): 746–758.

Xie, Y., Luu, K., and Savvides, M. (2012). A robust approach to facial ethnicity classification on large scale face databases. *Biometrics: Theory, Applications and Systems*, pages 143–149.

Xu, X., Tsang, I. W., and Xu, D. (2013). Soft margin multiple kernel learning. *IEEE Transactions on Neural Networks and Learning Systems*, 24(5):749–761.

Yang, M., Dai, D., Shen, L., and Gool, L. (2013a). Latent Dictionary Learning for Sparse Representation based Classification. *Proceedings of the IEEE*.

Yang, T., Thompson, P., Zhao, S., and Ye, J. (2017). Identifying Genetic Risk Factors via Sparse Group Lasso with Group Graph Structure. *arXiv preprint arXiv:1709.03645*.

Yang, M. and Moghaddam, B. (2000). Support vector machines for visual gender classification. In *Proceedings of the 15th International Conference on Pattern Recognition*, pages 1115–1118.

Yang, S., Shapiro, L., Cunningham, M., Speltz, M., Birgfeld, C., Atmosukarto, I., and Lee, S. I. (2013b). Skull retrieval for craniosynostosis using sparse logistic regression models. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7723 LNCS:33–44.

Yang, W., Sun, C., Zheng, W., and Ricanek, K. (2017). Gender classification using 3D statistical models. *Multimedia tools and applications*, 76(3), 4491–4503.

Yang, Z. and Ai, H. (2007). Demographic Classification with Local Binary Patterns. *International Conference on Biometrics*, pages 464–473.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zaharia, T. and Preteux, F. (2001). 3D-shape-based retrieval within the MPEG-7 framework. *Photonics West ...*.

Zech, W. D., Hatch, G., Siegenthaler, L., Thali, M. J., and Lösch, S. (2012). Sex determination from os sacrum by postmortem CT. *Forensic Science International*, 221(1-3):39–43.

Zhang, C. and Chen, T. (2001). Efficient feature extraction for 2D/3D objects in mesh representation. In *International Conference on Image Processing*, pages 935–938.

Zhang, G. and Wang, Y. (2009). Multimodal 2D and 3D Facial Ethnicity Classification. *2009 Fifth International Conference on Image and Graphics*, pages 928–932.

Zhang, L., João, M., and Ferreira, A. (2004). Survey on 3D shape descriptors. *DecorAR*, pages 1–28.

Zhao, P., Rocha, G., Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Annals of Statistics*, 37(6):3468–3497.

Zhu, X., Huang, Z., Cui, J., and Shen, H. T (2013). Video-to-shot tag propagation by graph sparse group lasso. *IEEE Transactions on Multimedia*, 15(3):633–646.

Zhou, Y, and Tuzel, O (2017). VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. *Arxiv*.