Facial Shape Space using Statistical Models from Surface Normals

Simone Regina Ceolin

Submitted for the degree of Doctor of Philosophy

University of York

Department of Computer Science

August 2011

Abstract

The analysis of shape-variations due to changes in facial expression and gender difference is a key problem in face recognition. In this thesis, we use statistical shape analysis to construct shape-spaces that span facial expressions and gender, and use the resulting shape-model to perform face recognition under varying expression and gender. Our novel contribution is to show how to construct shape-spaces over fields of surface normals rather than Cartesian landmark points. According to this model face needle-maps (or fields of surface normals) are points in a high-dimensional manifold referred to as a shape-space.

We start using a distance measure to discover the similarity between faces and gender difference, using a number of alternatives including geodesic, Euclidean and cosine distance between points on the manifold. For recognition we compare the perfomance distance between Euclidean, cosine and geodesic distance associated with the shape manifold. Here we explore if the distances used to distinguish gender and recognise the same for under different expressions. Also, to explore whether the Fisher-Rao metric can be used to characterise the shape changes due to differences expression and gender difference. We use a 2.5D representation based on facial surface normals (or facial needle-maps) for gender classification and face expression. The needle-map is a shape representation which can be acquired from 2D intensity images using shape-from-shading (SFS). Using the von-Mises Fisher distribution, we compute the elements of the Fisher information matrix, and use this to compute geodesic distance between fields of surface normals to construct a shape-space. We embed the fields of facial surface normals into a lower dimensional pattern space

using a number of alternative methods including multidimensional scaling, heat kernel embedding and commute time embedding. We present results on clustering the embedded faces using the BU-3DFEDB, Max Planck and EAR databases. The thesis is divided into five chapters.

In Chapter 1 we give a brief introduction and an outline of the thesis. Chapter 2 provides a literature review of aspects related to this topic. More specifically, it starts by describing the relevant literature of face expression recognition and gender difference in computer vision research and the approaches for facial shape recovery. Then, we give a detailed review of the methods developed for face expression and gender difference recognition. Also, we are working with facial shape recovery, we review some relevant techniques including the popular shape-from-shading methods. Are presented in Chapter 4 our first attempt to perform face expression recognition using facial needle maps which is recovered using shape-from-shading. We explore how the different distance measure can be used to measure different facets of shape including gender and expressions. We compute geodesic, euclidean, cosine and Mahalanobis distances for the long-vector representation of the needle-maps for the faces. In Chapter 5 we explore how the Fisher-Rao metric can be used to measure different facets of facial shape estimated from fields of surface normals using von-Mises Fisher (vMF) distribution. In particular we aim to characterise the shape changes due to differences in gender and due to different facial expression. We make use of the vMF distribution since we are dealing with surface normal data over the sphere R^2 The thesis concludes with Chapter 6, where a summary of the contributions of the research work is presented. Also, overall conclusions of the work, as well as, possible future work are discussed.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Edwin R. Hancock, whose guidance and support was paramount to this work. Also, I would like to thank my assessor,William Smith, for his support during the development of this work. I would also like to acknowledge ALBAN, one of the funding agencies of the Santander Group, which provided the financial support needed for this research work. Many thanks to all the members of the Computer Vision and Pattern Recognition Group, which provided a great working environment. A special thank you to Filo, who helped me from the first day until the last day. I would like to thank, with all my love, my parents, Tarcisio Ceolin, Vera Regina Achutti Cezar Ceolin, my sisters, Fabiana Ceolin and Cristiana Ceolin, my brother Tarcisio Ceolin Junior and the special children Gabriel, Clara and Nickiane. A special friend, Ana Cavalcanti, who I learned to have courage and believe in myself to finish my thesis. She is a true and very special friend that I got in England and my friend, Jim Woodcock, who gave me support and his knowledge to keep doing my thesis. To Ana and Jim my sincere thank you for everything and for the friendship. Also, Oliveira's family (Marcel, Lauana and Gabriela), Indrusiak's family (Leandro, Giuliana, Ana Laura and Lugui) and Dotti's family (Fernando, Leila, Luiza e Pedro) who provided many happy days and a friendship that will last forever. Friends who I would like to express my feeling for meeting them while in England, Leonardo, Clara, Andre, Alvaro, Artur, Oleg, Italo, Richard, Rosa and Ricardo. Finally, I would like to express my sincere gratitude to my husband, Dr. Osmar Marchi dos Santos. This thesis is special for you! I love you.

Declaration

I declare that the research work described in this thesis is original work unless otherwise indicated in the text. The research was conducted by me, Simone Regina Ceolin, between 2006 and 2011 at the University of York, with the supervision of Professor Edwin R. Hancock. Some parts of this thesis have been published in scientific publications.

List of Publications

Conference Papers

- Simone Ceolin, William A.P. Smith and Edwin R. Hancock. Facial Shape Spaces from Surface Normals and Geodesic Distance. IEEE Computer Society, pp. 416-423, 2007. The 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007).

- Simone Ceolin, William A.P. Smith and Edwin R. Hancock. Facial Shape Spaces from Surface Normals. Lecture Notes in Computer Science, pp. 955-965, v. 5112, 2008. Image Analysis and Recognition, 5th International Conference (ICIAR 2008).

- Simone Ceolin and Edwin R. Hancock. Using the Fisher-Rao Metric to Compute Facial Similarity. Lecture Notes in Computer Science, pp.384-393, v.6111, 2010. Image Analysis and Recognition, 7th International Conference (ICIAR 2010).

- Simone Ceolin and Edwin R. Hancock. Characterising facial gender difference using Fisher-Rao metric. 20th International Conference on Pattern Recognition (ICPR 2010). In Proc., pages 4308- 4311. IEEE.

- Simone Ceolin and Edwin R. Hancock. Distinguishing Facial Expression using the Fisher-Rao metric.17th International Conference on Image Processing (ICIP 2010). In Proc. pages 1437-1440.IEEE.

CHAPTER 1

Introduction

Face recognition is an extremely complex visual task. Over the past 30 years extensive research has been conducted by psychophysicists, neuroscientists, and engineers on various aspects of face recognition by humans and machines [125] [27] [88] [113]. Despite significant advances in face recognition, automated face recognition is the ultimate goal in the computer vision problem, with many commercial and law enforcement applications such as entertainment (e.g., for video games, virtual reality), smart cards (e.g., for drivers' licenses), information security (e.g., for personal device logon, database security) and law enforcement and surveillance (e.g., for CCTV control, suspect tracking and investigation) [100]. While the earliest work on face recognition can be traced back at least to the 1950s in psychology and to the 1960s

in the engineering literature. It is only very recently that acceptable results have been obtained. However, face recognition is still an area of active research since a completely successful approach or model has not been proposed to solve the face recognition problem.

These problems stem from the fact that existing systems are highly sensitive to environmental factors during image capture (Figure 1.1 shows examples of several variations), including: (i) Variations in facial orientation: the same face can be presented to the system at different perspectives and orientations. For instance, face images of the same person could be taken from frontal and profile views. Besides, head orientation may change due to translations and rotations. Rotation in depth and perspective transformation may also cause distortion. The common way to deal with pose variation is to isolate pose types (i.e., frontal, profile, rotated). (ii) Presence or absence of structural components: A robust face recognition system should be insensitive to noise generated by frame grabbers or cameras. Also, it should function under partially occluded images. A robust face recognition system should be capable of classifying a face image as "known" even under above conditions, if it has already been stored in the face database. Also, could be difficult to identify the gender. (iii) Facial expression: is a difficult task due to the natural variation in appearance of subjects. (iv) Illumination: face images of the same person can be taken under different illumination conditions such as, the position and the strength of the light source can be modified. Variations are the most difficult type to cope with due to fact that pixel intensities are directly affected in a nonlinear way by changing illumination intensity or direction.

This research focuses on the recognition of face expression and gender differ-

Figure 1.1: Examples of several variations during image capture.

ence because face expression and gender difference provide an important behavioral measure for the study of emotion, cognitive processes, and social interaction. Face expression and gender difference are commonly acceptable biometrics used by humans in their visual interaction. A biometric system is a pattern recognition system that establishes the authenticity of a specific physiological or behavioral characteristic.

The main motivation for this research topic is the analysis of shape-variations due to changes in facial expression and gender difference being a key problem in face recognition. Also, gender difference and face expression is an important issue for the development of intelligent systems. Some applications include making Human-Computer Interaction (HCI) more user-friendly, passive surveillance and access control, collecting valuable demographics (e.g., the number of women entering a store on a given day), and improving the performance of face identity recognition by using face expression specific models.

Statistical theories of shape variation [123] have been shown to be powerful tools for image interpretation. One important approach is to represent a shape by a set of landmark points on the boundary, and to capture shape variations using the covariance matrix for the Cartesian co-ordinates of the landmark points [31] [87]. Often, Cartesian landmarks are not the most convenient shape representation. For certain classes of shape, directional information is more convenient. However, if the statistical analysis of shapes is attempted with non-Cartesian data, then the construction of shape-spaces is no longer a straightforward problem.

This thesis aims to complement existing efforts in automatic face expression and gender difference recognition by taking into account the 3D facial information. The

shape information of a face can be encoded in a 3D head scan, a facial surface, or a facial needle-maps (a field of surface normals).

1.1 Motivation and Contribution

This work is motivated in part by the fact that faces have multiple shape properties, which can be used to categorize them according to different levels of specificity. Examples include gender, ethnicity, age, expression, identity, attractiveness and distinctiveness [123]. In particular, we are interested in how such shape variations manifest themselves in terms of changes in the field of surface normals. The reason for this is that we aim to fit statistical models of shape to 2D facial images, and from these images recover information concerning 3D shape. One natural way that captures features of the human vision system is to employ shape from shading to recover surface shape from variations in brightness. Here, it is more natural to represent the facial surface using fields of surface normals rather than surface height information, since the former is directly linked to the physical process of light reflectance.

We are particularly interested in the use of these ideas to represent variations in facial shape, and to determine the modes of variation due to factors such as gender and expression. Our contributions, following:

- Construct a shape-space that can be used to recognise instances of the same face from 2.5D images. As a result surface normal models are more suitable for the purposes of fitting to image data. However, due to their non-Cartesian nature the statistical modeling of variations in surface normal direction is more difficult than that for landmark positions. To overcome this problem, we

make use of the statistical representation developed by Smith and Hancock [88] which converts surface normals into a Cartesian form using the equidistance azimuthal projection. With the surface normal data in Cartesian form we construct a shape-space for variations due to changes in facial expression and gender difference.

- Construct a model for variations in facial shape due to changes in expression or different gender using information provided by facial needle-maps. A facial needle-maps is the description of the local orientation of a facial surface, from which the facial surface can be recovered by surface integration. Moreover, the orientation field is an intermediate representation in human visual perception, and is a component of the 2.5D sketch introduced by Marr [62]. This 2.5D representation is critical for shape-processing, and can be derived from 2D retinal images using shape-from-shading.

- Our goal in this thesis is to explore how to use the vMF distribution for shape representation using Fisher Information Matrix, and in particular to recognise variations in facial shape due to expression and gender difference. Fields of surface normals can be viewed as distributions of points residing on a unit sphere and may be specified in terms of the elevation and azimuth angles. It is natural to parameterise such statistical variations in direction using the von-Mises Fisher (vMF) distribution, which is specified in terms of a mean surface normal direction and a concentration parameter. Working in the surface normal domain (see Figure 1.2), we show how to use the vMF distribution to represent unstructured surface normal data without landmarks.

To measure the similarity between two fields of surface normals parameterized using the vMF distribution we make use of the Fisher-Rao metric. This way, facial similarity is measured by the geodesic distance between the shapes on a statistical manifold.

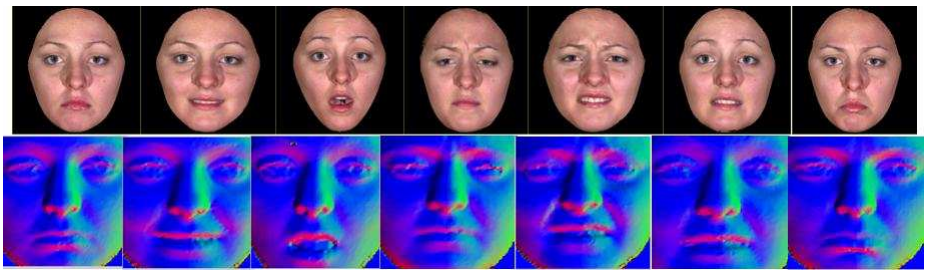

Figure 1.2: Face expression on fields of surface normals.

1.2 Thesis Outline

The remainder of this thesis is organized into the following chapters.

Chapter 2 provides a literature review of aspects related to the research topic. More specifically, it starts by describing the relevant literature of face expression and gender difference recognition in computer vision and the approaches for facial shape recovery. Then, we give a detailed review of the methods developed for face expression and gender difference recognition. Since we are also working with facial shape recovery, we review some relevant techniques including the popular shape-from-shading methods.

In Chapter 4 we present our first attempt to perform face expression recognition using facial needle maps, which is recovered using shape-from-shading. We explore

how different distance measures can be used to calculate different facets of shape including gender and expression. We compute geodesic, Euclidean, cosine and Mahalanobis distances for the long-vector representation of the needle-maps for the faces.

In Chapter 5 we explore how the Fisher-Rao metric can be used to measure different facets of facial shape estimated from fields of surface normals using the von-Mises Fisher distribution. In particular we aim to characterise the shape changes due to differences in gender and facial expressions. We make use of the vMF distribution since we are dealing with surface normal data over the sphere R^2.

The thesis concludes with Chapter 6, where a summary of the contributions of the research work is presented. Also, overall conclusions of the work, as well as, possible future work are discussed.

CHAPTER 2

Literature Review

In this thesis, we use statistical shape analysis to construct shape-spaces that span facial expressions and gender difference by facial needle-maps, and use the resulting shape-model to perform face recognition under varying expression and gender. Previous work have shown that the representation of shape has proved to be an elusive problem in the mathematics and statistics literature. This research involves three main themes: (i) Facial data representation. (ii) Measure distance to discover the similarity between face expressions and gender difference, which is measured using a number of alternative distances (geodesic, Euclidean, cosine and Mahalanobis). (iii) Explore whether Fisher-Rao metric can be used to characterise the shape changes due to expression and gender differences. In this chapter, we provide a thorough

review of the literature relevant to these topics. The chapter is organized as follows. We commence in Section 2.1 with a review about Face recognition including face expressions and gender difference. Section 2.2 shows how statistical classification could be represented. Section 2.3 provides the most popular statistical of shape used in the literature. In the section 2.4 shows how shape can be represented and we review the most popular statistical approaches to face recognition (face expression and gender). Section 2.5 describes geometric details about shape-space. Section 2.6 shows the most important statistical shape-space approaches and the most famous methods used by the research community. Finally, Section 2.7 concludes the chapter.

2.1 Face Recognition

An important goal in image understanding is to detect, track and label objects of interest present in observed images. Objects can be characterized in many ways according to their colors, textures, shapes, movements, and locations. The past decade has seen significant advances in the modeling and analysis of pixel values or textures to characterize objects in images with limited success. The study of human face recognition has remained a disparate field until 1975 [118] pointed out the lack of a broad theoretical framework to bring the findings together. Since then, other theories of face recognition have been proposed [12], [100], [18], [88], [106].

In this thesis we work with face expression and gender difference. There has been interest in both topics, face expression and gender difference, for scientists from several different fields such as computer science, engineering, psychology, neuroscience

and others [21] [27].

2.1.1 Face Expression

There are many ways that humans display their emotions. The most natural way to display emotions is using facial expressions (Figure 2.1). In the past 20 years there has been much research on recognizing emotion through facial expressions. The research about face expression is motivated because it is extensively used in a wide range commercial and law enforcement applications. This reasearch was pioneered by Ekman, Friesen and Ancoli [32] who started their work from the psychcology perspective. In the early 1990s the engineering community started to use these results to construct automatic methods of recognizing emotions from facial expressions in images or video [21] [63] [68] [79] [118] .

Work on recognition of emotions from image and video has been recently suggested and shown to work by Chen [21], Chen et al. [20], De Silva et al. [27], [88], [91]. These researches include two methods: holistic (Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)) and feature-based (Gabor and Scale Invariant Feature Transform-based methods). Holistic approaches use the entire face region for the task of feature extraction and, therefore, avoid difficulties in the detection of specific facial landmarks. Feature-based approaches on the other hand, extract local features from specific feature points of the face. Cohen [122] proposed a multilevel HMM that uses the state sequence of independent HMMs to segment and recognize facial expression.

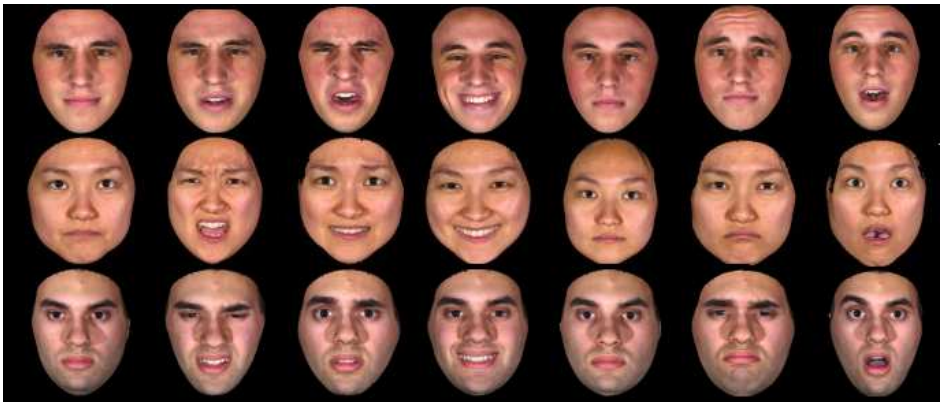

Figure 2.1: Face Expression - From left to right: Anger, Disgust, Fear, Happy, Neutral, Sad, Surprise.

2.1.2 Gender Difference

Women and men have on invariant aspect of faces (gender difference). Humans can readily determine gender difference for most faces (familiar or unfamiliar). Even with the hair style concealed, men's facial hair removed, and no cosmetic cues, human subjects can still determine sex with an accuracy as high as 96 percent [15]. This observation has attracted the interest of gender difference and prompted them to explore what is the facial information that differentiates males and females [14] [13]. Buchala [14] applied Principal Component Analysis (PCA) and explored the PCA components that gave the greatest gender discrimination using Linear Discriminant Analysis (LDA). Sun [94] used genetic algorithms to apply PCA features vectors to select the gender discriminating feature subset. Lu et al. [58] applied the use of a pixel-pattern-based texture feature for gender classification. The main idea of this method is that the face image can be regarded as a composition of micro-patterns. The pattern templates were obtained through PCA, and Adaboost was

used to select the discriminating feature subset. Also, Active Appearence Models (AAM) have been used by some researches as a feature extraction mechanism in gender discrimination. Saatci and Town [81] use AAM and Support vector machine for gender. Also, Wilhelm and Bohme [108] compare AAM with Independent Component Analysis (ICA) for gender difference. In [46], the differentiation capabilities of full face, jaw, lips/mouth, nose and eyes were evaluated. These regions were manually clipped, represented by AAM and classified using LDA. Shakhnarovich et al. [83] applied a thresholded weak classifier variant of Adaboost to detected face images for gender difference. Baluja et al. [2] explored the use of Adaboost on low resolution grayscale face images. Makinen and Raisamo [60] [59] combined face detection and gender classification, and gave a comprehensive comparison of state-of-art gender classification methods. Section 2.2.1 explain some statistical approaches to face recognition that were presented previously: Eigenfaces, Fisherfaces and AAM).

2.2 Statistical Classification

Statistical classification is represented in terms of d features or measurements viewed as a point in a d-dimensional space. The goal is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a dimensional feature space [43].

Statistical pattern recognition has been used successfully to design a number of commercial recognition systems. In statistical pattern recognition, a pattern is represented by a set of d features, or attributes, viewed as a d-dimensional feature

vector.

One of the most important stages in statistical face recognition is dimensionality reduction to represent each face by a d-dimensional feature vector, where d is as small as possible. There are some reasons for the low-dimensional representation.

The first reason is that no matter if the original data representation is $2D$ intensity images, or $2\frac{1}{2}D$ surfaces, or $3D$ models, it will inevitably contain information which is either redundant or irrelevant to the face attribute recognition task. According to Marr [62], the processes involved in vision produce a series of representations providing increasingly detailed information about the visual environment. Marr identified three major kinds of representation [62], [109]:

- (i) $2D$: this provide a two-dimensional description of the main light-intensity changes in the visual input, including information about edges and contours. This representation is obtained from the two-dimensional intensity images;
- (ii) $2\frac{1}{2}$D sketch: this incorporates a description of the depth and orientation of visible surfaces. Using the technology of Shape-From-Shading, this representation can be achieved from a single intensity image. Like the primal sketch, it is observer-centered;
- (iii) $3D$ model representation: this describes three-dimensionally the shapes of objects and their relative positions in a way that is independent of the observer's viewpoint.

The second reason as stated in [29] is that, the classification of patterns as performed by humans is based on very few of the most important features. Moreover,

keeping the number of features to the absolute minimum will curtail the effect of the curse of dimensionality phenomenon on the complexity of the classifier. Mathematical dimensionality reduction techniques can be classified into two major categories:

- (a) Feature selection: select the best subset of the input features in the measurement space. Feature selection is more useful in the face attributes (e.g., gender, ethnicity, age) recognition rather than in the face identity recognition, because very few components are required to encode attributes such as gender, ethnicity and age. However, a large number of components are required to encode identity and these components are widely distributed [14];
- (b) Feature extraction create new features in a transformed space. Popular feature extraction methods include Principal Component Analysis (PCA) [44], Principal Geodesic Analysis (PGA) [35], Linear Discriminant Analysis (LDA) [95].

2.2.1 Statistical Approaches to Face Recognition

Since the 1990s, appearance based methods have been dominant, from which two face recognition methods are derived: holistic appearance feature based, and analytic local feature based. In recent years, Eigenfaces [106] and Fisherfaces [3] approaches have attracted much attention.

2.2.1.1 Eigenfaces

Eigenfaces was proposed by Turk and Pentland for face identification in 1991 [106]. In their work, face images are projected into a feature space (face space) through

Principal Component Analysis [70] [42]. PCA is a technique that can used to simplify a data set. The PCA aim to reduce the dimensionality of the data set and to identify new meaningful underlying variables. PCA is widely used in the statistical face recognition area. The most famous is the Eigenfaces [106] [74]. The Eigenfaces are the principal component of the original face image, obtained by the decomposition of PCA, forming the face space from these images.

Consider a set of N sample images $x_1, x_2, ...x_N$ taking values in an dimensional image space, and assume that each image belongs to one of c classes $X_1, X_2, ...X_c$. Consider a linear transformation mapping the original dimensional image space into n-dimensional feature space, where $m < n$. The new feature vectors $y_k \in R^m$ are defined by the following linear transformation:

$$y_k = W^T x_k \tag{2.1}$$

$$k = 1, 2, ..., N \tag{2.2}$$

where $W \in R^{n \times m}$ is a matrix with orthonormal columns. If the total scatter matrix S_T is defined as

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T. \tag{2.3}$$

where $\mu \in R^n$ is the mean image of all samples, then after applying the linear transformation W^T, the scatter of the transformed feature vectors $(y_1, y_2, ..., y_N)$ is $W^T S_T W$.

In [99], who observed that the conventional definition of PCA lacked an associ-

ated probabilistic model for the observed data. PCA can be derived within a density estimation framework, and proposed their probabilistic Principal Component Analysis method. Based on the probabilistic framework of PCA, they further proposed a mixture of probabilistic principal component analyzer [98], which is able to model complex data structures with a combination of local PCA models. Moghaddam and Pentland [67] also combined density estimation with eigenspace decomposition, and proposed a probabilistic visual learning technique. Moreover, they showed how the learning technique was effectively applied to face detection and recognition, especially to the view-based face recognition.

Scholkopf et al. [82] proposed the kernel PCA method. By making use of kernel functions to compute the dot products between two vectors, they efficiently computed principal components in a high-dimensional feature space, which is related to the input space by a non-linear mapping. In [119] and [51], the kernel PCA method has been applied to face recognition and they achieved a positive performance.

Koren et al. [53] have proposed a weighted PCA method. They investigated the relationship between PCA and multidimensional scaling, and considered PCA as computing the projection that maximizes the sum of squared distances in the projected feature space. By incorporating symmetric non-negative pairwise weights (dissimilarities) into the sum of squared distances, they controlled the directions of the leading PCA eigenvectors.

2.2.1.2 Fisherfaces

Fisherfaces was proposed by Belhumeur et al. [3] in 1997. It is based on Fisher Linear Discriminant Analysis (LDA). LDA is a widely used method for feature

extraction and dimensionality reduction in pattern recognition and it tries to find the best project direction in which training samples belonging to different classes are best separated.

This method selects W in such a way that the ratio of both between class scatter and the within class scatter is maximized. Let us consider a set of N sample images $(x_1, x_2, ..., x_N)$ taking values in an n-dimensional image space, and assume that each image belongs to one of c classes $(X_1, X_2, ..., X_c)$. Let the between class scatter matrix defined as

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T. \tag{2.4}$$

and the within class scatter matrix be defined as

$$S_T = \sum_{k=1}^{N} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T. \tag{2.5}$$

where μ_i is the mean image of class X_i, N_i is the number of samples in class X_i, and μ is the mean image of all samples.

Zhao et al. proposed a subspace LDA method to solve the generalized and overfitting problems when performing face recognition on a large face data set but with very few training face images available per class. This method consists of two steps. First, the face image is projected into a face subspace via PCA. Second, LDA is applied to the PCA projected vectors to construct a linear classifier in the subspace. They also employed a weighted distance metric guided by the LDA eigenvalues to improve the performance of face recognition using the subspace LDA method.

2.3 Statistical Shape Analysis

Statistical shape analysis involves methods for the geometrical study of random objects where location, rotation and scale information can be removed. The subject is a new and exciting area of statistics, offering many fresh challenges. Recently, many research have beeen using shape for analysing in many fields such as computer vision, medical imaging, pattern recognition and others ([24] [26] [91] [41]).

Previous work has shown that the representation of shape has proved to be an elusive problem in the mathematics and statistics literature. The development of a rigorous statistical theory of shape began with the work by Kendall [49] which describes the shape formed by a set of random points under Brownian motion, and has been used in the statistical analysis of shape in archaeology and astronomy. Bookstein [10] and Ziezold [126], on the other hand, have developed methods for analysing the variations in biological forms.

In the image analysis literature there are numerous examples of the use of Kendall's shape spaces [48], [87] and [102]. Developments in statistical shape theory due to Small [86] and [87] suggest that improved Shape Spaces can be obtained by representing objects as points on a high-dimensional surface (a manifold) in such a way that different views of a given object correspond to a single point on the manifold.

In recent years, several authors have proposed shape prior segmentation algorithms [26], [55], [80]. The general idea is always to include some a priori information about the shape in the segmentation algorithm in order to limit the space of possible solutions.

The value of using shape priors has been shown in a variety of tracking contexts. Model-based methods [4] have been used in computer vision for a long time, especially for rigid objects. Deformable templates [122] are an effective and powerful method to model prior shapes and allow for many deformations modes of shapes.

However, the modelling of objects by such methods generally requires many parameters and is often done heuristically. Dynamical models, which implicitly contain shape priors, are extensively used in object tracking to improve the robustness by estimating state spaces and predicting possible movements [96].

The design of shape priors depends on ongoing work on statistical shape models [23], [31] and [47]. Statistical models of shape variation were proposed by Wang and Staib [107], Cremers et al. [25], and Rousson and Paragios [80]. Leventon et al. [55] defined a statistical shape model to guide the evolution process.

Chen et al. [22] introduced a shape prior segmentation method using the curve representing the boundary of an object as the shape representation. The method introduces a shape prior in the geodesic active contour framework proposed by Caselles [16]. Rousson and Paragios [80] used a signed distance function as a shape representation.

Several authors have used this shape representation in order to introduce the shape priors in their proposals [26], [105]. Paragios et al. [80] use a probabilistic level set distance map. Other shape distances have also been used, such as the Hausdorff distance in [19].

In particular, advanced models of shape spaces, shape distances, and corresponding shape transformations have been proposed recently [19], [52], [37], [84] and [121].

To construct the shape prior model a shape representation is selected and then

all of the training instances are coded with that shape representation.

2.4 Shape Representation

The field of geometrical shape analysis was initially developed from a biological point of view by D'arcy Thompson [97] and is a field of increasing importance. Shape analysis is of great importance in a wide variety of disciplines, e.g., image segmentation, image understanding, classification and many other fields.

Shape is commonly used in an everyday language sense, usually referring to the appearance of an object. An intuitive definition of shape is given by Kendall [49]. Following Kendall [47], the definition of shape is *"All the geometrical information that remains when location, scale, and rotational effects are filtered out from an object"*. So, an object's shape is invariant under the Euclidean similarity transformations of translation, scaling, and rotation.

Two objects have the same size-and-shape if they can be translated and rotated to each other so that they match exactly, i.e., if the objects are rigid-body transformations of each other. Definition of size and shape: two geometric objects have the same size and shape, if one can be mapped onto the other by an Euclidian (rotation and translation) transformation.

Also, we can describe shape by locating a finite number of points on each specimen, which are called landmarks. When analysing shape, it is important that the landmarks correspond throughout the training set. *Landmark is a point of correspondence on each object that matches between and within population.* There are three basic types of landmarks: anatomical, mathematical and pseudo-landmarks

[31]. (i) Anatomical landmarks: are points that correspond between samples in some biologically meaningful way, e.g., the tip of a finger. Often anatomical landmarks are assigned by experts. Letting an expert annotate each of a series of images with a set of corresponding anatomical landmarks is a direct and certain method to generate a training set. In many cases this is the best method, but it is also the most expensive and time consuming method. (ii) Mathematical landmarks: are points located on an object according to some mathematical or geometrical property of the boundary, e.g., at a point of high curvature or at an extreme point. The use of mathematical landmarks is particularly useful in automatic recognition and analysis. If two dimensional points could be placed at clear corners of object boundaries or 'T' junctions between boundaries. (iii) Pseudo-landmarks: there are seldom enough mathematical and/or anatomical landmarks to give more than a sparse description of the shape of the target object. One way to obtain a richer shape description is to place pseudo-landmarks in between the well-defined landmarks. Pseudo-landmarks are points along the boundaries which are arranged to be uniformly spaced between well-defined landmark points.

2.5 Shape Space

In this section, we further investigate geometrical aspects of shape. The shape of a geometric object in $\mathbb{R}^n$ is represented by a set of k landmarks, each of which is a point in $\mathbb{R}^n$. In practice $n = 2$ or $n = 3$.

For computation, a shape defined by landmarks is also represented by a vector

$$x = \begin{bmatrix} l^1 \\ \vdots \\ l^k \end{bmatrix}$$

by column stacking the landmarks. The vector is called a shape vector.

A landmark l is in general represented by a $n \times 1$ vector and the shape vector x becomes naturally a $nk \times 1$ vector. In two dimensions a landmark can be represented by a complex number and x is in C^k.

Each training set of annotated objects $X1, ..., X_n s$, can be represented by a matrix $X = [x1, ..., x_n s]$, where x_i denotes the shape vector corresponding to the annotated object $X_i (i = 1, ..., n_s)$.

Quantitative study of shape begins with the formulation of a shape space, in which each shape is represented as a point. A distance metric on this shape space gives a measure of shape similarity between any two objects.

Pre-Shape space is the space of all possible pre-shapes. Formally, the pre-shape space S_m^k is the orbit space of the non-coincident k point set configuration in R^m under the action of translation and isotropic scaling. The term pre-shape means that we are one step from shape-rotation still has to be removed.

In the shape space of Figure 2.2 we can see that different views of a given object correspond to single point. In a shape space of $2D$ objects, all possible views of an object caused by translation, scaling and rotation are represented as a single point and the object recognition can then be achieved, for example, by computing the geodesic distance between the object and a model in the shape space.

The idea of a shape-space $\sum_m^K$, whose elements are the shapes of k labelled

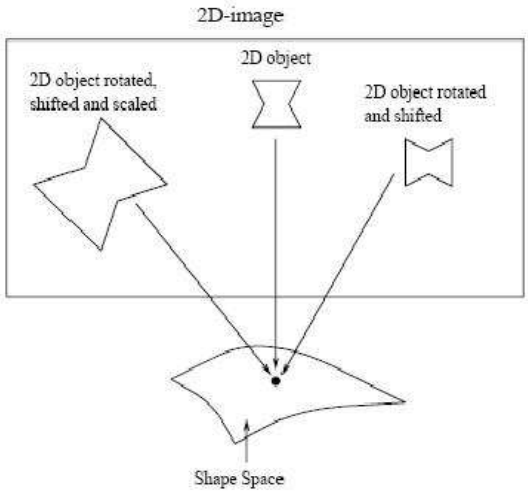

Figure 2.2: Illustration of a shape space for $2D$ objects [123].

points in R^m, at least two being distinct, was introduced in a statistical context by Kendall (1984) [48].

Here, it is natural to identify shapes differing only by translations, rotations, and scaling in R^m (although there are situations of interest, not to be considered here, in which the scale is also relevant). However, this identification will not apply to reflections. Thus, the resulting shape space is the quotient

$$\sum_{m}^{k} = \frac{S^{m(k-1)-1}}{SO(m)}. \tag{2.6}$$

of the sphere by the rotation group. This is the base space of a fibre bundle with total space $S^{m(k-1)-1}$ and fibre SO(m). The former is naturally endowed with a

uniform Riemannian metric, while the latter, being a compact Lie group, possesses an invariant metric.

The key observation of Kendall is that the metrical geometry of these quotient spaces, long studied by geometers, is precisely the required tool for the introduction of measures appropriate for the systematic comparison and classification of various shapes.

Now, for a planar distribution of points, with $m = 2$, the shape space $\sum_2^k$ of k points is simply a complex projective space P^{k-2} of dimension $k - 2$, with the Fubini-Study metric defining the geodesic distances between pairs of shapes.

The vertices of a shape in R^2 (viewed as a complex plane), with the centroid at the origin, determine the homogeneous coordinates of a point in P^{k-2}. The permutation group $\sum(k)$ of order k interchanges homogeneous coordinates, and is therefore represented by projective unitary transformations of P^{k-2}.

For three-point configurations, i.e., $k = 3$, the shape space is a complex projective line P^1, which, viewed as a real manifold, is a two-sphere S^2. Hence, the natural shape space for triangles is essentially a Riemann sphere (Kendall 1985, Watson 1986).

2.6 Statistical Shape Spaces Approaches

In this section, a brief overview about statistical shape spaces approaches is presented. More details about this topic, can be found at [52], [92], [91], [49] and [39]. There are three different approaches to the study of shapes: (i) Landmark-Based Shape Analysis, (ii) Geometric Analysis of shapes of curves and (iii) Diffeo-

morphisms for shape matching. In this section we show a brief description about Landmark-Based Shape Analysis since it is used in this work.

2.6.1 Landmark-Based Shape Analysis

The first formal mathematical theory of shapes is due to David Kendall [49] and is referred to as a Procrustes shape analysis. A remarkable body of work on analysis of shapes also exists due to works of Bookstein ([10], [8] and [7]), Mardia ([61]), Small ([87], [86] and [54]), and others. These studies share the property that objects (and their shapes) are represented using a finite number of landmarks (points in Euclidean spaces) and equivalences are established with respect to shape preserving transformations. Landmarks are chosen manually to capture shapes of objects in images. There have been at least three distinct origins of what we call shape theory. The first approach seems to have been that of Kendall ([49]) who was, at the time, concerned with shape in archaeology and astronomy. Kendall ([48] and [50]) described shape spaces for comparing point sets in any number of dimensions. While the space is directly related to Procrustes's distance, it undergoes a Riemannian submersion to create a non-Euclidean metric. Distances between point sets are then measured as geodesics along the surface of a hyper-sphere. We refer to these distances as Kendall's distance (Kdist). For point sets in three or more dimensions, the space contains singularities

$$\cos^2(Kdist) = \frac{(\sum_i Z_i^A \overline{Z}_i^B)(\sum_i \overline{Z}_i^A Z_i^B)}{(\sum_i Z_i^A \overline{Z}_i^A)(\sum_i Z_i^B \overline{Z}_i^B)}. \tag{2.7}$$

where $\overline{Z}$ is the complex conjugate of Z, $\overline{Z} = x - iy$. Before comparing two points

sets, they are corrected for translation by placing their centroids at zero by setting.

There is an emerging family of techniques for studying the shapes of connected curves in Euclidean spaces. Both the geometry of curves and the geometry of spaces of curves play a role in this shape analysis. Srivastava [52] proposed differential geometric representations of curves using their direction functions and curvature functions. Shapes are represented as elements of infinite-dimensional spaces and their pairwise differences are quantified using the lengths of geodesics connecting them on these spaces. They use a Fourier basis to represent tangents to the shape spaces and then use a gradient-based shooting method to solve for the tangent that connects any two shapes via a geodesic.

There are many interesting applications of the shape theory proposed by [52]. An important advantage of this approach is that it provides geodesic paths between arbitrary shapes on the shape spaces. These paths can be used to interpolate between shapes, extrapolate a shape change, and compute a mean shape under a probability distribution on shapes.

2.7 Conclusions

It should be clear from this literature review that (1) extensive research in shape analysis has been conducted during the last few decades, and (2) there are many shortcomings in the current state of the art. We conclude this chapter with a summary of the above literature, and state the position of this thesis in the context of the broad field.

Our main objective in this thesis is to use a 2.5D representation of facial shape

(facial needle-maps) which can be recovered from 2D facial images. Using a novel face representation for face expression and gender difference to construct shape-space, this thesis fil a gap in the face expression and gender difference literature. The first step of our technique is to use surface normal data in Cartesian form to construct a shape-space for variations due to changes in facial expression and gender difference.

A research about geometric shape-space and the most important statistical shape-space approaches, and most famous methods used by the research community were reviewed. In the literature review, we see a gap on the use of geometric shape-space to recognise face expression and gender difference from surface normals. As like described in the literature review, geometric shape space techniques are used to recognise objects and other types of recognition, which do not include face expressions and gender difference ([52], [75] [24]).

After comparing the studies about shape space approaches, we aim to extend the shape-space developed by Small [87] from the domain of Cartesian landmark points to a field of surface normals and explore how the Fisher-Rao metric can be used to measure different facets of facial shape estimated from fields of surface normals using von-Mises Fisher distribution. In particular we aim to characterise the shape changes due to differences in gender and due to different facial expression. We make use of the vMF distribution since we are dealing with surface normal data over the sphere R^2.

CHAPTER 3

Cartesian Representation of Vector Field

In this section, we explain how directional data can be converted into a Cartesian form using the exponential map from a manifold to a tangent space.

A unit vector $\mathbf{n} \in \mathbb{R}^3$ may be considered as a point lying on a spherical manifold $n \in S^2$, where S^2 is the unit 2-sphere. The two are related by $\mathbf{n} = \Phi(n)$ where $\Phi : S^2 \mapsto \mathbb{R}^3$ is an embedding. Likewise, a field of N surface normals $\mathbf{U} \in R^{N\times 3}$ describing a surface may be considered as a point on a manifold $U \in S^2(N) = \prod_{i=1}^{N} S^2$.

3.0.1 The Log and Exponential Maps

If $v \in T_n S^2$ is a vector on the tangent plane to S^2 at $n \in S^2$ and $v \neq 0$, the *exponential map*, denoted Exp_n, of v is the point on S^2 along the geodesic in the direction of v at distance $\|v\|$ from n. Geometrically, this is equivalent to marking out a length equal to $\|v\|$ along the geodesic that passes through n in the direction of v. The point on S^2 thus obtained is denoted Exp_n^v. This is illustrated in Fig. 4.1.

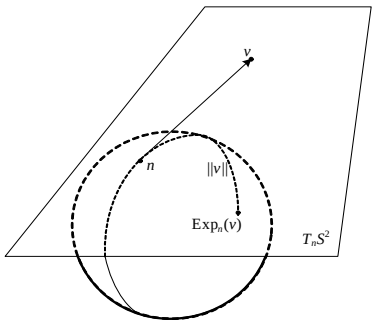

Figure 3.1: The exponential map.

The inverse of the exponential map is the *log map*, denoted Log_n. Therefore, the equality $\text{Log}_n(\text{Exp}_n(v)) = v$ holds. The geodesic distance between two points $n_1 \in S^2$ and $n_2 \in S^2$ can be expressed in terms of the log map, i.e. $d(n_1, n_2) = \|\text{Log}_{n_1}(n_2)\|$. The exponential and log maps for the space of a field of N surface normals, $S^2(N)$, are simply the direct products of N copies of the maps for S^2 given above.

3.0.2 Spherical Medians and Variance

A distribution of spherical directional data $\mathbf{n}_1, \ldots, \mathbf{n}_K \in \mathbb{R}^3$ can be characterised using the mean direction [61] $\hat{\mathbf{n}}_0 = \frac{\bar{\mathbf{n}}}{|\bar{\mathbf{n}}|}$ where $\bar{\mathbf{n}} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{n}_i$. If we consider

the distribution of unit vectors as a distribution of points on a spherical manifold $n_1, \ldots, n_K \in S^2$, where $\Phi(n_k) = \mathbf{n}_k$, it is clear that the mean direction is dependent on the embedding Φ and is the *extrinsic mean* of a distribution of spherical data: $\mu_\Phi = \arg\min_{n \in S^2} \sum_{i=1}^{K} \| \Phi(n) - \Phi(n_i) \|^2$.

If we define the projection $\pi : \mathbb{R}^3 \mapsto S^2$ as $\pi(\mathbf{n}) = \arg\min_{n \in S^2} \|\Phi(n) - \mathbf{n}\|^2$, We may show that the mean direction is the extrinsic mean:

$\mu_\Phi = \pi(\bar{\mathbf{n}}) = \pi\left(\frac{1}{K}\sum_{i=1}^{K} \Phi(n_i)\right)$.

In other words, the extrinsic mean is the Euclidian average (or centre of mass) of the distribution of points in $\mathbb{R}^3$, projected back onto the closest point on the sphere. A more natural definition of the average of a distribution of points on the unit sphere uses arc length as the choice of distance measure. Since a 2-sphere is a Riemannian manifold and great circles are geodesics, arc length is the Riemannian distance $d(.,.)$ between a pair of points and hence, $d(n_1, n_2) = \arccos(\Phi(n_1) \cdot \Phi(n_2))$. Using this definition of distance, we can define the *intrinsic mean*: $\mu = \arg\min_{n \in S^2} \sum_{i=1}^{K} d(n, n_i)$.

For spherical data, this is known as the *spherical median* [34]. This point cannot be found analytically, but can be solved iteratively using the gradient descent method of Pennec [72]. We initialise our estimate as the Euclidian mean of distribution, i.e. $\mu_{(0)} = \mu_\Phi$. The current estimate is then updated iteratively as follows:

$$\mu_{(j+1)} = \mathrm{Exp}_{\mu_{(j)}}\left(\frac{1}{K}\sum_{i=1}^{K} \mathrm{Log}_{\mu_{(j)}}(n_i)\right). \tag{3.1}$$

To find the intrinsic mean $\mu \in S^2(N)$ of a sample of K fields of N surface normals: $U_1, \ldots, U_K \in S^2(N)$, we replace the exponential and log maps in Equation 4.2 with the corresponding maps for the space $S^2(N)$. We can use the log map

and intrinsic mean to define the sample variance of a distribution of points on the sphere: $\sigma_2 = \frac{1}{K}\sum_{i=1}^{K} d(\mu, n_i)^2 = \frac{1}{K}\sum_{i=1}^{K} \|\mathrm{Log}_{\mu}(n_i)\|^2$.

Suppose that each of the K training examples is a range image which consists of an array of depth data each containing $N = X_{\mathrm{res}} \times Y_{\mathrm{res}}$ pixels. For the pixel indexed p in the k_{th} training sample the depth is z_p^k. Using the range data we estimate the surface normal directions, and the surface normal at the pixel location p for the k_{th} training image is $\mathbf{n}_p^k$.

We calculate the spherical median μ_p of the distribution of surface normals $\mathbf{n}_p^1, \ldots, \mathbf{n}_p^K$ at each pixel location p using (4.2). The surface normal $\mathbf{n}_p^k$ is represented by its position on the tangent plane $T_{\mu_p}S^2$ given by the log map: $v_p^k = \mathrm{Log}_{\mu_p}(\mathbf{n}_p^k) \in \mathbb{R}^2$

A field of surface normals projected to the tangent plane to their local spherical median may be represented as the long vector: $\mathbf{v}^k = [v_1^k, \ldots, v_N^k]^T$

With the intrinsic mean of the distribution to hand, we can transform each field of surface normals representing a facial surface to a distribution of 2-dimensional points in a Cartesian space using the log map. These projected points retain their variance with respect to the average direction and provide a convenient representation with which to work.

CHAPTER 4

Distances to measure different facets of facial shape

Face recognition has gained much attention in recent years and has become one of the most successful applications of image analysis and understanding. Face recognition conferences are emerging and sophisticated commercial systems have been developed that achieve rather high recognition rates. There is an important goal in image analysis to classify and recognise objects of interest in given image and imaged objects can be characterized in several ways, using their colors, textures, shapes, movements and locations. However, analysis of shapes, expecially complex objects (faces), is a challenging task and requires sophisticated mathematical tools. Most of the researches are based on 2D intensity images on the recognition of identities of people. Also, few studies have been investigated the role of 3D facial shape in gender

difference and face expression due to the more complex computations required in 3D sensors currently available.

Contributions

This chapter shows our first attempt to perform gender difference and face expression needle-maps using Shape-from-Shading (SFS). We present a 2.5D data set. This chapter presented the preliminary results. The aim of our experiments is to explore how the different distance measures can be used to measure different facets of facial shape including gender and expressions. We compute geodesic, Euclidean, Mahalanobis and cosine distances for the long-vector representations of the needle-maps of faces. We visualise the distribution of distances using multidimensional scaling [54] to embed the faces in a two-dimensional pattern space (details of MDS are given in Section 4.2).

For the recovery of facial shapes from 2D images, we have a special interest in statistical model based SFS and geometric SFS. SFS methods satisfying statistical model constraints [30] [88] [89] [90] guarantee that the recovered shapes lie in a space representing realistic facial shapes, and therefore avoid invalid recoveries such as a concaved nose. Geometric SFS methods [17] [88] [90] recover facial shapes satisfying data closeness as a hard constraint. As a result, the recovered shapes implicitly encode the intensity information. This is especially useful for recognition tasks. Principal geodesic shape-from-shading combines the statistical constraint and data-closeness constraint, and is therefore capable of recovering valid and accurate facial shapes. Compared to the popular 3D morphable model [6], principal geodesic shape-from-shading is more efficient and improves the model dominance. In this thesis,

we choose principal geodesic shape-from-shading to recover facial needle-maps from 2D images for gender difference and face expression.

To construct the statistical model, we make use of Principal Geodesic Analysis (PGA) [35] [73]. For the analysis of facial needle-maps, PGA is better suited since a needle-map is a field of surface normals which reside on a spherical manifold. After the recovery of facial needle-maps, we use PGA to find the principal geodesics that capture the largest statistical variance in the recovered facial needle-maps. Then, each facial needle-map is represented by a parameter vector, referred to as PGA feature vector. Each component in the feature vector tells the projected position of this needle-map on the corresponding principal geodesic.

The remainder of this chapter is organized as follows. Section 4.1 describes how directional data can be converted into a Cartesian form using the exponential map from a manifold to a tangent space and describes how to incorporate the statistical model in shape-from-shading. Also, Section 4.2 describes embedding techniques to embed the facial shapes into a two-dimensional pattern space. The method studied is Multi-Dimensional Scaling (MDS) and we compare the results with the two other embeddings systems: Heat kernel and Commute Time. Next, in Section 4.3 we show how Small's ideas can be used to construct a shape-space for data in this form. In Section 4.5, we first give a description of the data set and the normalization method, then we show the results achieved using the four distance measures (Euclidean, geodesic, cosine and Mahalanobis) applied for the recognition of gender and expressions. Finally, Section 4.6 concludes this chapter.

4.1 Facial Shape Recovery

We recover facial needle-maps using an SFS method, referred to as Principal Geodesic SFS (PGSFS). The idea is to augment shape-from-shading using a statistical model that captures variations present in fields of surface normals extracted from human-faces. In this section, we describe how directional data can be converted into a Cartesian form using the exponential map from a manifold to a tangent space, and how to apply Principal Geodesic Analysis (PGA) to a set of example facial needle-maps for the purpose of learning a statistical model of facial shape.

4.1.1 Cartesian Representation of Vector Field

In this section, we explain how directional data can be converted into a Cartesian form using the exponential map from a manifold to a tangent space.

A unit vector $\mathbf{n} \in \mathbb{R}^3$ may be considered as a point lying on a spherical manifold $n \in S^2$, where S^2 is the unit 2-sphere. The two are related by $\mathbf{n} = \Phi(n)$ where $\Phi : S^2 \mapsto \mathbb{R}^3$ is an embedding. Likewise, a field of N surface normals $\mathbf{U} \in R^{N\times 3}$ describing a surface may be considered as a point on a manifold $U \in S^2(N) = \prod_{i=1}^{N} S^2$, which is a special case of a non-linear Riemmanian manifold. To model the distribution of data on this manifold, we use the Principal Geodesic Analysis (PGA), which is a generalization of Principal Component Analysis (PCA) from Euclidean data to data residing on a Riemmanian manifold. PGA makes use of the intrinsic means of the data on the manifold and the Lie group representation based on log and exponential maps.

4.1.2 The Log and Exponential Maps

If $v \in T_nS^2$ is a vector on the tangent plane to S^2 at $n \in S^2$ and $v \neq 0$, the *exponential map*, denoted $\exp_n$, of v is the point on S^2 along the geodesic in the direction of v at distance $\|v\|$ from n. Geometrically, this is equivalent to marking out a length equal to $\|v\|$ along the geodesic that passes through n in the direction of v. The point on S^2 thus obtained is denoted $\exp_n^v$. This is illustrated in Figure 4.1.

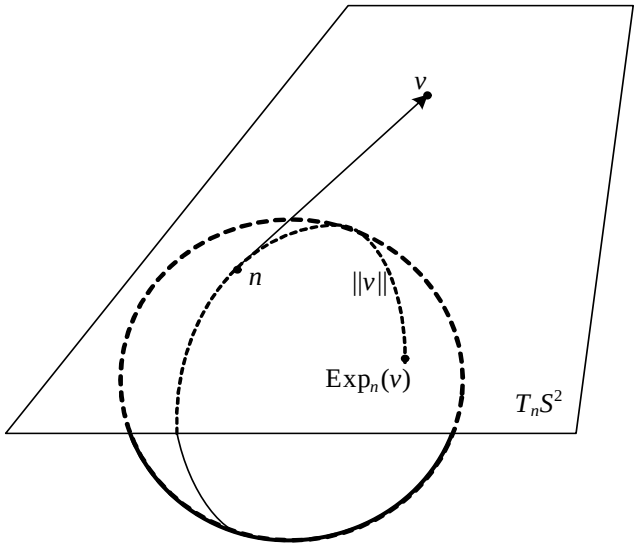

Figure 4.1: The exponential map.

The inverse of the exponential map is the log *map*, denoted $\log_n$. Therefore, the equality $\log_n(\exp_n(v)) = v$ holds. The geodesic distance between two points $n_1 \in S^2$ and $n_2 \in S^2$ can be expressed in terms of the log map, i.e., $d(n_1, n_2) = \|\log_{n_1}(n_2)\|$. The exponential and log maps for the space of a field of N surface normals, $S^2(N)$, are simply the direct products of N copies of the maps for S^2 given above.

4.1.3 Spherical Medians and Variance

A distribution of spherical directional data $\mathbf{n}_1, \ldots, \mathbf{n}_K \in \mathbb{R}^3$ can be characterised using the mean direction [61] $\hat{\mathbf{n}}_0 = \frac{\bar{\mathbf{n}}}{|\bar{\mathbf{n}}|}$ where $\bar{\mathbf{n}} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{n}_i$. If we consider the distribution of unit vectors as a distribution of points on a spherical manifold $n_1, \ldots, n_K \in S^2$, where $\Phi(n_k) = \mathbf{n}_k$, it is clear that the mean direction is dependent on the embedding Φ and is the *extrinsic mean* of a distribution of spherical data: $\mu_\Phi = \underset{n\in S^2}{\arg\min} \sum_{i=1}^{K} \| \Phi(n) - \Phi(n_i) \|^2$.

If we define the projection $\pi : \mathbb{R}^3 \mapsto S^2$ as $\pi(\mathbf{n}) = \underset{n\in S^2}{\arg\min} \|\Phi(n) - \mathbf{n}\|^2$, we may show that the mean direction is the extrinsic mean: $\mu_\Phi = \pi(\bar{\mathbf{n}}) = \pi\left(\frac{1}{K}\sum_{i=1}^{K}\Phi(n_i)\right)$.

In other words, the extrinsic mean is the Euclidian average (or centre of mass) of the distribution of points in $\mathbb{R}^3$, projected back onto the closest point on the sphere. A more natural definition of the average of a distribution of points on the unit sphere uses arc length as the choice of distance measure. Since a 2-sphere is a Riemannian manifold and great circles are geodesics, arc length is the Riemannian distance $d(.,.)$ between a pair of points and hence, $d(n_1, n_2) = \arccos\left(\Phi(n_1)\cdot\Phi(n_2)\right)$. Using this definition of distance, we can define the *intrinsic mean*:

$$\mu = \underset{n\in S^2}{\arg\min} \sum_{i=1}^{K} d(n, n_i). \tag{4.1}$$

For spherical data, this is known as the *spherical median* [34]. This point cannot be found analytically, but can be solved iteratively using the gradient descent method of Pennec [72]. We initialise our estimate as the Euclidian mean of distribution, i.e., $\mu_{(0)} = \mu_\Phi$. The current estimate is then updated iteratively as follows:

$$\mu_{(j+1)} = \exp_{\mu_{(j)}} \left(\frac{1}{K} \sum_{i=1}^{K} \log_{\mu_{(j)}}(n_i) \right). \tag{4.2}$$

To find the intrinsic mean $\mu \in S^2(N)$ of a sample of K fields of N surface normals: $U_1, \ldots, U_K \in S^2(N)$, we replace the exponential and log maps in Equation 4.2 with the corresponding maps for the space $S^2(N)$. We can use the log map and intrinsic mean to define the sample variance of a distribution of points on the sphere: $\sigma_2 = \frac{1}{K} \sum_{i=1}^{K} d(\mu, n_i)^2 = \frac{1}{K} \sum_{i=1}^{K} \|\log_{\mu}(n_i)\|^2$.

Suppose that each of the K training examples is a range image which consists of an array of depth data each containing $N = X_{\text{res}} \times Y_{\text{res}}$ pixels. For the pixel indexed p in the k_{th} training sample the depth is z_p^k. Using the range data we estimate the surface normal directions, and the surface normal at the pixel location p for the k_{th} training image is $\mathbf{n}_p^k$.

We calculate the spherical median μ_p of the distribution of surface normals $\mathbf{n}_p^1, \ldots, \mathbf{n}_p^K$ at each pixel location p using Equation (4.2). The surface normal $\mathbf{n}_p^k$ is represented by its position on the tangent plane $T_{\mu_p} S^2$ given by the log map: $v_p^k = \log_{\mu_p}(\mathbf{n}_p^k) \in \mathbb{R}^2$.

A field of surface normals projected to the tangent plane to their local spherical median may be represented as the long vector: $\mathbf{v}^k = [v_1^k, \ldots, v_N^k]^T$.

With the intrinsic mean of the distribution to hand, we can transform each field of surface normals representing a facial surface to a distribution of 2-dimensional points in a Cartesian space using the log map. These projected points retain their variance with respect to the average direction and provide a convenient representation with which to work.

4.1.4 Principal Geodesic Analysis for facial Needle-maps

In this section, we explain how to apply PGA to a set of example facial needle-maps for the purpose of learning a statistical model of facial shape.

PGA is a generalization of PCA from data residing in a Euclidean space to data residing on a Riemmanian manifold. The goal of PCA is to find a linear subspace of the space in which the data lies, and maximize the variance of the projected data in the subspace. In PGA, the notion of a linear subspace is replaced by that of a geodesic manifold. The geodesics that traverse the submanifold are referred to as principal geodesics. They are analogous to the principal axes in PCA, expect that each principal axis in PCA is a straight line. In the spherical case, a principal geodesic corresponds to a great circle.

To project a point $n_1 \in S^2$ onto a great circle C is to find the point on C that is nearest to n_1 in terms of geodesic distance. The projection $\pi_c : S^2 \rightarrow C$ is defined as: $\pi_C(n_1) = \arg\min_{n \in S^2} d(n1, n)^2$ where $d(n_1, n)$ is the geodesic distance between n_1 and n on the spherical manifold. For a geodesic C passing through the intrinsic mean μ, this projection can be approximated linearly in the tangent space $T_\mu S^2$

$$\log_\mu(\phi_C(n_1)) \approx \sum_{i=1}^{d} v_i . \log_\mu(n_1), \tag{4.3}$$

where $v_1, ..., v_d$ is an orthonormal basis for $T_\mu S^2$, and can be obtained using standard PCA. Then, the principal geodesic for the S^2 space are obtained under the exponential map $\exp_\mu(v_1), i = 1...d$. This approximation enables us to compute the principal geodesics by applying PCA in the tangent plane $T_\mu S^2$.

Suppose there are K example facial needle-maps, each having N pixel locations.

The surface normal at the pixel location l for k^{th} needle-map is n_{kl}. The intrinsic mean μ_l of the surface normals $n_{1l}, ..., n_{kl}$ at each pixel location l is calculated. The surface normal n_{kl} is then represented by its log mapped position $u_{kl} = \log n_{kl}$ in the tangent plane $T_\mu S^2$. The process is illustrate in Figure 4.2.

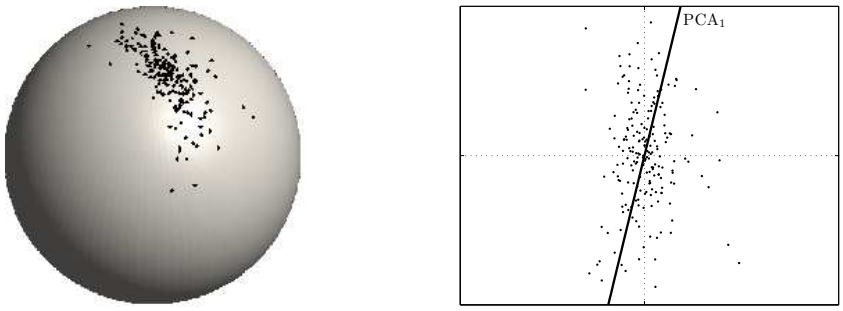

Figure 4.2: Projection of surface normals on the unit sphere (a) to points on the tangent plane at the mean (b) [88].

On the right are the log mapped positions of the points with the mean as the center of the projection. For the k^{th} training needle-map, we concatenate the x, y-coordinates of u_{kl} at the N pixel locations, and form the $2N$ dimensional log mapped long vector $u_k = [u_{k1x}, u_{k1y}, \ldots, u_{kNx}, u_{kNy}]^T$ in the tangent plane $T_\mu S^2(N)$. The K long vectors form the column-wise data matrix $U = [u_1|\ldots|u_K]$, and the covariance matrix is $\Sigma = \frac{1}{K}UU^T$.

Because N, the dimensionality of the facial needle-maps, is usually too large to make the manipulation of Σ feasible, the numerically efficient snap-shot method of Sirovich [85] is used to compute the eigenvectors of Σ. Accordingly, we construct the matrix $\hat{\Sigma} = \frac{1}{K}U^TU$, and locate its eigenvalues and eigenvectors. The i^{th} eigenvector e_i of Σ can be computed from the i^{th} eigenvector $\hat{e}_i$ of $\hat{\Sigma}$ using $e_i = U\hat{e}_i$. The i^{th} eigenvalue λ_i of Σ equals the i^{th} eigenvalue $\hat{\lambda}_i$ of $\hat{\Sigma}$ when $i \leq K$. When $i > K$,

$\lambda_i = 0$. The $K-1$ leading eigenvectors of Σ form the columns of the eigenvector matrix (projection matrix) $\Phi = (e_1|e_2|\ldots|e_{K-1})$, where K is the number of sample facial needle-maps. Given a facial needle-map, the log mapped long vector $u = [u_{1x}, u_{1y}, \ldots, u_{Nx}, u_{Ny}]^T$ is computed, then the corresponding PGA feature vector is $b = \Phi^T u$. From the PGA feature vector $b = [b_1, \ldots b_{K-1}]^T$, the needle-map can be generated using: $n_l = \exp_{\mu_l}((\Phi b)_l)$ at each location l.

4.1.5 Incorporating the Statistical Model into SFS

With the mean μ and projection matrix Φ of the statistical model constructed using PGA from the sample facial needle-maps, we are able to augment shape-from-shading using this model in an iterative way. This shape-from-shading method is referred to as the principal geodesic shape-from-shading (PGSFS). It not only imposes the image irradiance equation as a hard local brightness constraint [111], but also satisfies a strict global shape constraint by projecting the recovered surface normals into the space spanned by the constructed statistical model. The steps of this SFS method are illustrated in Figure 4.3. The facial needle-map is initialized as the intrinsic mean of the statistical model $n^{(0)} = \mu$. Since the statistical model captures the distribution of facial surface normals, the projection into the model space guarantees that the recovered needle-map represents a valid human face.

Satisfying the statistical constraint

Given an estimated facial needle-map $n^{(t)}$, we can first compute its log mapped long vector $u^{(t)} = \log_\mu n^{(t)}$, and then obtain its best fit PGA parameters $b^{(t)} = \Phi^T u^{(t)}$. Next, the updated facial needle-map, which is an instance of the statistical model,

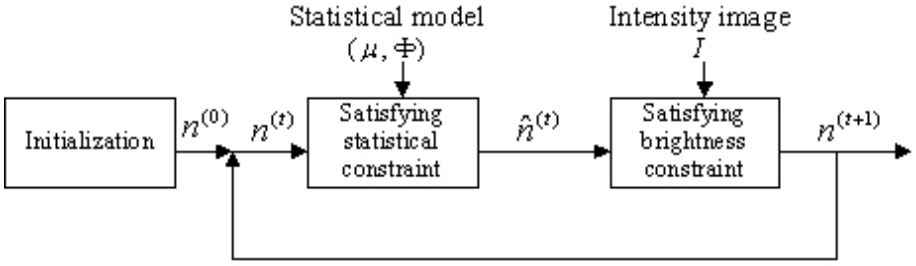

Figure 4.3: Process of PGSFS [112].

can be generated as $\hat{n}^{(t)} = \exp_{\mu}(\Phi^{(t)})$. Because the statistical model captures the variance structure of facial needle-maps, by satisfying the statistical constraint, the generated facial needle-map guarantees that it has a realistic appearance. This also overcomes the well-known local convexity-concavity instability problem [69] in previous SFS methods.

Satisfying the brightness constraint

Let $I_l \in R$ denote the intensity at the pixel location l. According to Worthington and Hancock [111], when the surface reflectance follows Lambert's law $I_l = n_l \cdot s$, then the surface normal is constrained to fall on a cone whose axis is in the light source direction s and whose opening angle is $\alpha = \arccos I_l$. With $\hat{n}^{(t)}$ generated from the statistical model at hand, in order to enforce this brightness constraint, we rotate the normal at each pixel location l back to its closest on-cone position. This is equivalent to moving each surface normal $\hat{n}_l^{(t)} \in S^2$ along a passing geodesic from $s \in S^2$ a distance $\arccos I_l$. This is illustrated in Figure 4.4 which shows the tangent plane $T_s S^2$. The updated surface normal at pixel location l is

$$n_l^{(t+1)} = \exp_s \left\{ \arccos(I_l) \frac{\log_s(\hat{n}_l^{(t)})}{\log_s(\hat{n}_l^{(t)})\|} \right\}.$$

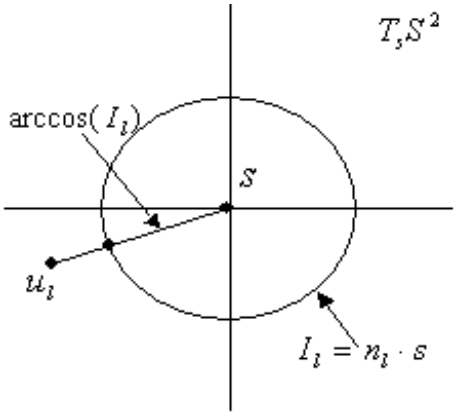

Figure 4.4: Restoring a normal to the closest position on the reflectance cone [90].

Since the facial needle-map satisfies the brightness constraint as a hard constraint, they effectively encode the original image data. In other words, the original image can be recovered from the needle-map. As a result, the needle-map implicitly encodes image intensity.

Algorithm

The PGSFS algorithm can be summarized as follows [112]:

1. Initialize: $n^{(0)} = \mu$, where μ is the intrinsic mean of the model. Set iteration $t = 0$.

2. Estimate PGA parameters: $b^{(t)} = \Phi^T \log_\mu(n^{(t)})$. Generate best fit normals: $\hat{n}^{(t)} = Exp_\mu(\Phi b^{(t)})$.

3. Update normals at each pixel location l, $l = 1, \ldots, N$:

$$n_l^{(t+1)} = \exp_s \left\{ \arccos(I_l) \frac{\log_s(\hat{n_l}^{(t)})}{\log_s(\hat{n_l}^{(t)})\|} \right\},$$

Set $t = t + 1$.

4. Stop if $t > max$ iterations.

Upon convergence, there are two recovered needle-maps, a) $\hat{n}$ is an instance of the statistical model, referred to as the best fit needle-map, and b) n is the closest on-cone position of $\hat{n}$. Therefore n not only satisfies data-closeness constraint, but also satisfies the statistical model which guarantees that the recovered shape realistically resembles a human face.

Given a statistical model constructed by applying PGA to a set of facial needle-maps, the recovered facial needle-map n can be represented by the model parameters $b = \Phi^T \log_\mu(n)$, where Φ is the projection matrix of the model, and μ is the model mean. b is referred to as the PGA feature vector of the recovered facial needle-map n. The statistical model can be the one used in PGSFS, which is constructed from range data. It can also be constructed from a set of recovered facial needle-maps. In this chapter, we adopt the latter, and represent the recovered facial needle-maps by their PGA feature vectors obtained using the model constructed from themselves.

4.2 Embedding Techniques

To visualise the distribution of geodesic distances we use a number of manifold embedding techniques to embed the facial shapes into a two-dimensional pattern space. The method studied is Multi-Dimensional Scaling (MDS) [54]. We compare the results with the other two embedding systems: Heat kernel [1] and Commute Time [40].

4.2.1 Multi-Dimensional Scaling

Multi-Dimensional Scaling (MDS) [54] is a procedure which allows data specified in terms of a matrix of pairwise distances to be embedded in a Euclidean space. The classical multi-dimensional scaling method was proposed by Torgenson [101] and Gower [38]. Here we intend to use this method to embed shock trees in a low-dimensional space.

The first step to perform the classical MDS is to compute the squared distance matrix

$$D = [d_{i,j}^2], i, j = 1, ..., n. \tag{4.4}$$

where

$$d_{i,j}^2 = (x_i - x_j)^T (x_i - x_j). \tag{4.5}$$

Then, the inner product matrix is given by $B = -\frac{1}{2}JDJ$, where $J = \frac{I^{-1}}{n^{11^T}}$ and $L_1 \times n = [l, l, ..., l]$. B is symmetric, positive semi-definite and of rank K which is the original dimension of n points. Therefore B has k non-negative eigenvalues and $n - k$ zero eigenvalues.

The matrix B can be expressed in terms of its spectral decomposition, $B = V\Lambda V^T$, where $\Lambda_n \times n = diag(\lambda_1, \lambda_2, \lambda_K, 0, ..., 0)$. For convenience, the eigenvalues of B are ordered such that $\lambda_1 \geq \lambda_2 \geq ...\lambda_K \geq 0$. Hence, the required coordinates are given by using the non-zeros sub matrix $\Lambda_{K\times K}$, $X = V_1\Lambda_1^{\frac{1}{2}}$, where V_1 and Λ_1 are the first p eigenvectors and eigenvalues of B.

4.2.2 Heat Kernel

Heat Kernel commence from Laplacian matrix, following the work of Bai and Hancock [114] and El-Ghawalby and Hancock [33]. Suppose that the graph under study is denoted by $G = (V, E)$, where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. The adjacency matrix A for the nodes is:

$$A(u,v) = \begin{cases} 1 & if (u,v) \in E \\ 0 & otherwise \end{cases} \tag{4.6}$$

The Laplacian matrix is $L = D - A$, where D is the diagonal degree matrix with the elements $D(u,u) = \sum_{v \in V} A(u,v)$. The normalized Laplacian is given by $\hat{L} = D^{\frac{-1}{2}} L D^{\frac{-1}{2}}$. The spectral decomposition of the normalized Laplacian matrix is:

$$\hat{L} = \Phi \Lambda \Phi^T = \sum_{i=1}^{|V|} \lambda_i \phi_i \phi_i^T. \tag{4.7}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \lambda_|V|)$ is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\phi_1|\phi_2|....|\phi_|V|)$ is the matrix with the ordered eigenvectors as columns. Since $\hat{L}$ is symmetric and positive semi-definite, the eigenvalues of the normalized Laplacian fall in the interval $[0, 2]$, i.e., they are all positive. The eigenvector associated with the smallest non-zero eigenvector is referred to as the Fielder-vector. Using the heat equation associated with the Laplacian,

$$\frac{\partial h_t}{\partial_t} = \hat{L} h_t. \tag{4.8}$$

where h_t is the heat kernel and t is the time. The solution is found by exponentiating the Laplacian eigenspectrum, i.e.,

$$h_t = \sum_{i=1}^{|V|} \exp[\lambda_i t]\phi_i\phi_i^T = \Phi \exp[-t\Lambda]\Phi^T. \tag{4.9}$$

The heat kernel is a $|V| \times |V|$ matrix, and for the nodes u and v of the graph G the resulting component is

$$h_t(u,v) = \sum_{i=1}^{|V|} \exp[-\lambda_i T]\phi_i(u)\phi_i(v). \tag{4.10}$$

When t tends to zero, then $h_t \cong I - \hat{L}t$, i.e., the kernel depends on the local connectivity structure or topology of the graph. If, on the other hand, t is large, the $h_t \cong \exp[-t\lambda_m]\phi_m\phi_m^T$, where λ_m is the smallest non-zero eigenvalue and ϕ_m is the associated eignvector, i.e., the Field vector. Hence, the large time behaviour is governed by the global structure of the graph. It is interesting to note that the heat kernel is also related to the path length distribution on the graph. If $P_k(u,v)$ is the number of paths of length K between nodes u and v then

$$h_t(u,v) = \exp[-t] \sum_{k=1}^{|V|^2} P_k(u,v)\frac{t^k}{k!}. \tag{4.11}$$

The path-length distribution is itself related to the eigenspectrum of the Laplacian. By equation 4.11 derivatives of the spectral and path-length forms of the heat kernel show that

$$P_k(u,v) = \sum_{i=1}^{|V|} (1-\lambda_i)^k \phi_i(u)\phi(v). \tag{4.12}$$

When the graph is embedded on a manifold in Riemannian space then the pattern of geodesic distances between nodes on the manifold is the same as the path length distribution. However, when the manifold is locally Euclidean, then the heat kernel is approximated by the Gaussian

$$h_t(u,v) = [4\phi]^{\frac{-n}{2}} \exp[\frac{-1}{4t^2}d(u,v)^2]. \tag{4.13}$$

where d(u,v) is the distance between the nodes u and v on the Euclidean manifold and n is the dimensionality of the space. The aim here is to find an approximation to the geodesic distance between nodes in the embedding, by equating the spectral and Gaussian forms for the Kernel. The result is

$$d(u,v) = \sqrt[2]{-tln\bigg((4\pi t)^{\frac{n}{2}} \sum_{i=1}^{|V|} \exp[-\lambda_i t]\phi_i(u)\phi_i(v)\bigg)}. \tag{4.14}$$

We can consider the behaviour of this function for large and small values of t. When t is small, making use of the fact that $h_t = I - \hat{L}t$, we have

$$d(u,v) = \sqrt[2]{-t\bigg(\frac{n}{2}ln[4\pi t] + ln[1 - \hat{L}(u,v)t\bigg)}. \tag{4.15}$$

Hence, the small t behaviour is determined by the local topology of the graph. Moreover, since the second term under the square-root vanishes, the behaviour near $t = 0$ is independent of the structure of the graph. On the other hand, when t is large, we can write

$$d(u,v) = \sqrt[2]{-t\bigg(\frac{n}{2}ln[4\pi t] - \lambda_m t + ln\phi_m(u)\phi_m(v)\bigg)}. \tag{4.16}$$

For very large t we have that $d(u,v) \cong t\sqrt{\lambda_m}$ and hence the effect of local edge-structure is completely smoothed away. Although the parameter t potentially provides a route to a graph scale-space, here the $4\pi t$ is constant $= 1$.

4.2.3 Commute Time

The commute-time between nodes a and b is the expected number of steps for a discrete-time random walk to reach node b from a and then return again. The embedding which preserves commute time as Euclidean distance between nodes has the co-ordinate matrix $Y_C = \frac{Vol}{\sqrt{\Lambda_-}}\Phi_-^T$ where $Vol = \sum_{a=1}^{N} D(a,a)$ is the volume and Λ_- and Φ_- are matrices obtained by deleting the rows and columns corresponding to zero elements of the Laplacian eigenvalue matrix.

4.3 Four different distance measures

The representation of shape has proved to be an elusive problem in the mathematics and statistics literature. The development of a rigorous statistical theory of shape began with the work of Kendall [49] which describes the shape formed by a set of random points under Brownian motion, and has been used in the statistical analysis of shape in archaeology and astronomy. Bookstein [10] and Ziezold [126], on the other hand, have developed methods for analysing the variations in biological forms. In the image analysis literature there are numerous examples of the use of Kendall's Shape Spaces [48], [87].

Recent developments in statistical shape theory due to Small [86] and [87] suggest that improved shape spaces can be obtained by representing objects as points on a

high-dimensional surface (a manifold) in such a way that different views of a given object correspond to a single point on the manifold.

In Kendall's [48] work, the shape of the point configuration $x_1, ..., x_k$ is represented as an element on a shape manifold $\sum_n^k$ whose Riemannian geometry is motivated by Procrustes arguments. A consequence of this geometric representation of shape is that the similarity of shapes can be measured by the geodesic distance on the manifold between the shapes corresponding to any two configurations of landmarks.

Small's [87] idea is to combine Kendall's geometry theory of shape with embedding methods based on multidimensional scaling. The starting point for the embedding process is a dissimilarity matrix.

If X_j, $j = 1, ..., m$, is a set of m such $k-$landmark configurations, the $m \times m$ dissimilarity matrix has an element in the row i and column j which is the geodesic distance between the shape configuration X_i and the shape of configuration X_j in $\sum_n^k$ [87].

Suppose that $x_1, ..., x_{n+1}$ points in general position in $\mathbb{R}^n$. We let $X = (x_{ij})$ be the $n \times (n+1)$ matrix whose j_{th} column is the vector x_j. Let $Y = (y_k l)$ denote this $n \times n$ matrices of full rank. The pre-shape matrices can be placed into $1-1$ correspondence with $GL(n)$, the General Linear Group on $\mathbb{R}^n$.

The $n \times n$ matrix Y formed by these columns of vector differences can then be multiplied by an orthogonal $n \times n$ matrix ϑ on the left so that the resulting product ϑY. We say $Z = \vartheta Y$ is a version of the original configuration matrix standardized both with respect to location and orientation [87].

Small [87] refers to $Z = (z_{kl})$ as the form-matrix of the configuration matrix

X. Let $T_+(n)$ be the class of $n \times n$ upper triangular matrices of positive diagonal elements. If we denote the group of orthogonal transformations of $\mathbb{R}^n$ by $O(n)$, then $O(n)$ is a subgroup of $GL(n)$. The natural space in which the form matrix Z resides is identifiable as the factor group $GL(n)/O(n)$. For any $Y \in GL(n)$, there is a unique factorization $Y = \vartheta Z$ where $\vartheta \in O(n)$ and $Z \in T_+(n)$. $T_+(n)$ is homeomorphic with the orbit space $GL(n)/O(n)$.

The quotient map from $GL(n)/O(n)$ is denoted by π. So, in terms of the above notation, we have $\pi(Y) = Z$. The final stage of the reduction is from the form-matrix Z to the shape matrix W. If we scale the elements of the form Z so that the scaled matrix has determinant 1, then scale information is eliminated from the form-matrix, leaving a matrix containing only information concerning the shape of the configuration of landmarks. Let $W = (w_{kl})$ be this scaled matrix.

The matrix W is an element of the set of the all upper triangular matrices with positive diagonal elements whose product is one. Let $ST_+(n)$ be the class of all such matrices. To obtain a natural metric on $ST_+(n)$, we begin by noting that the 2-form $tr[dYY^{-1}(dYY^{-1})^t]$ on $GL(n)$ is both positive definite and symmetric and so provides a right-invariant metric tensor on $GL(n)$. Since the elements of $O(n)$ acting on the left are isometries with respect to this metric, there is an induced differential structure and metric tensor on the orbit space $T_+(n)$, with respect to which the quotient map $\pi : GL(n) \mapsto T_+(n)$ is a Riemannian submersion. Moreover the matrix dYY_{-1} of 1-form on $GL(n)$ is right-invariant in the sense that, for any fixed Y_0 in $GL(n)$, the right multiplication $y \mapsto \check{Y} = YY_0$ gives $d\check{Y}\check{Y}^{-1} = dYY^{-1}$.

Having obtained a metric for $GL(n)$ in the formula $tr[dYY^{-1}(dYY^{-1})^t]$ a metric for $SL(n)$ is immediately obtained by restriction, because $SL(n)$ is submanifold

of $GL(n)$. It also possible to use formula $tr[dYY^{-1}(dYY^{-1})^t]$ and the quotient map $\pi : SL(n) \mapsto S_+(n)$: the induced metric tensor on $ST_+(n)$ is the horizontal component of the metric tensor on $SL(n)$, thinking of the fibers of π as vertical. The details of this construction can be found in [87].

It can shown that $ST_+(n)$ has constant negative curvature -2, and is isometric with the Poincaré Half Plane. However, when $n > 2$, although $ST_+(n)$ has non-positive curvature, it does not have constant curvature.

Now, to compute the geodesic distance between different shapes consider $GL(n)$ first. Since the metric tensor determined by the formula $tr[dYY^{-1}(dYY^{-1})^t]$ is right-invariant, the geodesic distance between any two points Y_1 and Y_2 is the same as that between I and $Y = y_2 y_1^{-1}$.

So the induced metric on $T_+(n)$ is also right-invariant, in the sense that, for any Y_0 in $GL(n)$, $dg(\pi(Y_1), \pi(Y_2)) = dg(\pi(Y_1 Y_0), \pi(Y_2 Y_0))$, where dg is the geodesic distance on $T_+(n)$. For $\pi(Y) \in ST_+(n)$, Le and Small [54] have shown that

$$dg(\pi(Y_1), \pi(Y_2)) = \sqrt{\sum_{i=1}^{n} \left[\log \left(\frac{\gamma_i}{(\prod_{j=1}^{n} \gamma_j)^{\frac{1}{n}}} \right) \right]^2}. \tag{4.17}$$

In this thesis, we construct a shape-space for variations due to changes in facial expression and gender difference. The similarity between faces is measured by the geodesic distance on the shape manifold using information provided by surface normals. For the computation of the geodesic distance on shape-space we have implemented the method proposed by Small [87] when $n = 2$: $dg(\pi(Y_1), \pi(Y_2)) = \frac{1}{\sqrt{2}} \left| \ln \frac{\gamma_2}{\gamma_1} \right|$, where γ_i are the positive square roots of the eigenvalues of $(Y_2 Y_1^{-1} - Y_2 Y_1^{-1})$. More detail about Small's shape space construction can be found in [87].

We will compare the performance of the geodesic distance with that of the Euclidean, cosine and Mahalanobis distances. For the long-vectors v_1^k and v_2^k the Euclidean distance is represented by the equation

$$d_e^2 = (v_1^k - v_2^k)^T (v_1^k - v_2^k). \tag{4.18}$$

and the cosine distance is represented by the equation

$$d_c = \arccos\left(\frac{|v_1^k| \cdot |v_2^k|}{||v_1^k|| \times ||v_2^k||}\right). \tag{4.19}$$

and the Mahalanobis distance is

$$d_m = \sqrt{(v_1^k - v_2^k)^T P^{-1} (v_1^k - v_2^k)}. \tag{4.20}$$

where P is the covariance matrix.

4.4 Overview of Data Sets

In this research we use several data sets for executing the experiments work, found in the literature.

4.4.1 Gender Data Sets

For the gender experiments, two different data sets are used. The first data set is from Notre Dame biometric database (EAR) [36] [18] consists of range images(3D) and frontal facial images (2D) selected from the University of Notre Dame Biometrics

database (Data set 2) [1].

The second data set is from the Max-Planck face database [6] [104] . It comprises 200 laser scanned (Cyberware TM) human heads (100 females and 100 males) which have smooth and precise facial surfaces. The facial needle-maps are obtaneid by first orthographically projecting the facial range scans onto a frontal view plane, and then aligning the plane according to the eye centers, and cropping the plane 142×124 pixels to maintain only the inner part of the face. Finally, the surface normal at each pixel position is computed using gradients of the processed range image.

4.4.2 Face Expression Data Set

For face recognition we are using the 3DFEDB database [56]. The database presently contains 71 subjects (females and males), with ages ranging ages from 18 years to 70 years old, with a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino. Each subject performed seven expressions in front of the 3D face scanner. With the exception of the neutral expression, each of the six prototypic expressions include: happiness, disgust, fear, angry, surprise and sadness.

4.4.3 Data Set and Normalization

For the experiments we use the Notre Dame Biometrics database described in section 4.4.1. In particular, before using the Notre Dame data set ([36], [18]) for gender

[1]University of Notre-Dame biometrics database. The database is available at http://www.nd.edu

discrimination, the range of images and the 2D facial image for each subject must be aligned and normalized.

4.4.4 Geometric Normalization

Geometric normalization is required for the range images which consist of height values z sampled at different image locations (x, z). Seven points are manually selected from the face (two points in each eye, one point in the nose, one point in the middle of the mouth and one point in the chin). We use the following references for the points: left eye (inside and the outside corners - $1, 2$), right eye (inside and the outside corners - $3, 4$), the nose (N), the middle of the mouth (M), and the center of the chin (C). The centers of the left and right eyes (denoted as Le and Re) are calculated as the midpoints of points 1 and 2, and 3 and 4 respectively. Firstly,we rotate and translate the range images so that the plane passing through Le, Re and C is perpendicular to the $Z - axis$, the line passing through Le and Re are horizontal, and the (x, z) position of N is $(0, 0)$.

The rotation matrix M_r is define in [57]:

$$M_r = M_z.M_x.M_y \tag{4.21}$$

where:

$$M_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix}$$

$$M_y = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}$$

$$M_z = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\alpha = arctan\left(\frac{y_0}{\sqrt{x_0^2 + z_0^2}}\right). \tag{4.22}$$

$$\beta = arctan\left(\frac{x_0}{z_0}\right). \tag{4.23}$$

$$\gamma = arctan\left(\frac{Le_y - Lr_y}{Le_x - Lr_x}\right). \tag{4.24}$$

where $(x_0, y_0, z_0) = (\vec{Le} - \vec{C}) \times (\vec{Lr} - \vec{C})$. If $N1$ denotes the position of N after rotation, the translation matrix is define as:

$$m_t = (-N1_x, -N1_y, 0)^T. \tag{4.25}$$

After rotation and translation, the point (x,y,z) becomes

$$(x', y', z')^T = M_r.(x, y, z)^T + M_t. \tag{4.26}$$

We calculate the mean positions of the five points Le', Lr', N', M' and C' from

the 200 range images, and use them as the reference points. Then, we first scale the range images to make the distance between Le' and Lr' identical to the reference. Then the principal warps method described in [9] is used to warp the range images so that the (x, y) positions of the five points (Le', Lr', N', M' and C') are identical to those of the reference ones. The nose N' gives the centerline for cropping a 114×100 region from the warped range images, which maintains only the inner part of faces. Linear interpolation is then used to fill the holes. The geometric normalization for $2D$ images is almost identical to that used for the $3D$ range images, except that the rotation and cropping are performed in the XY plane only.

4.4.5 Brightness Normalization

In addition to the geometric normalization, brightness normalization is also required for the $2D$ images. Firstly, the color images are converted into greyscale by averaging the values of the three color channels. The intensity contrast is then linearly stretched to normalize the ambient lighting variations using the formula:

$$I_{norm(x,y)} = \frac{I_(x, y) - I_{min}}{I_{max} - I_{min}}. \tag{4.27}$$

where $I(x, y)$ denotes the intensity value at location (x, y), I_{min} and I_{max} are the minimum and maximum intensity values in the image. Finally, we use the method proposed in [78] to apply photometric correction and specularity subtraction to the stretched intensity images in order to improve the results of PGSFS (which relies on the Lambartian reflectance model). After normalization, we can calculate facial needle-maps using the range images. We use the facial needle-maps extracted from

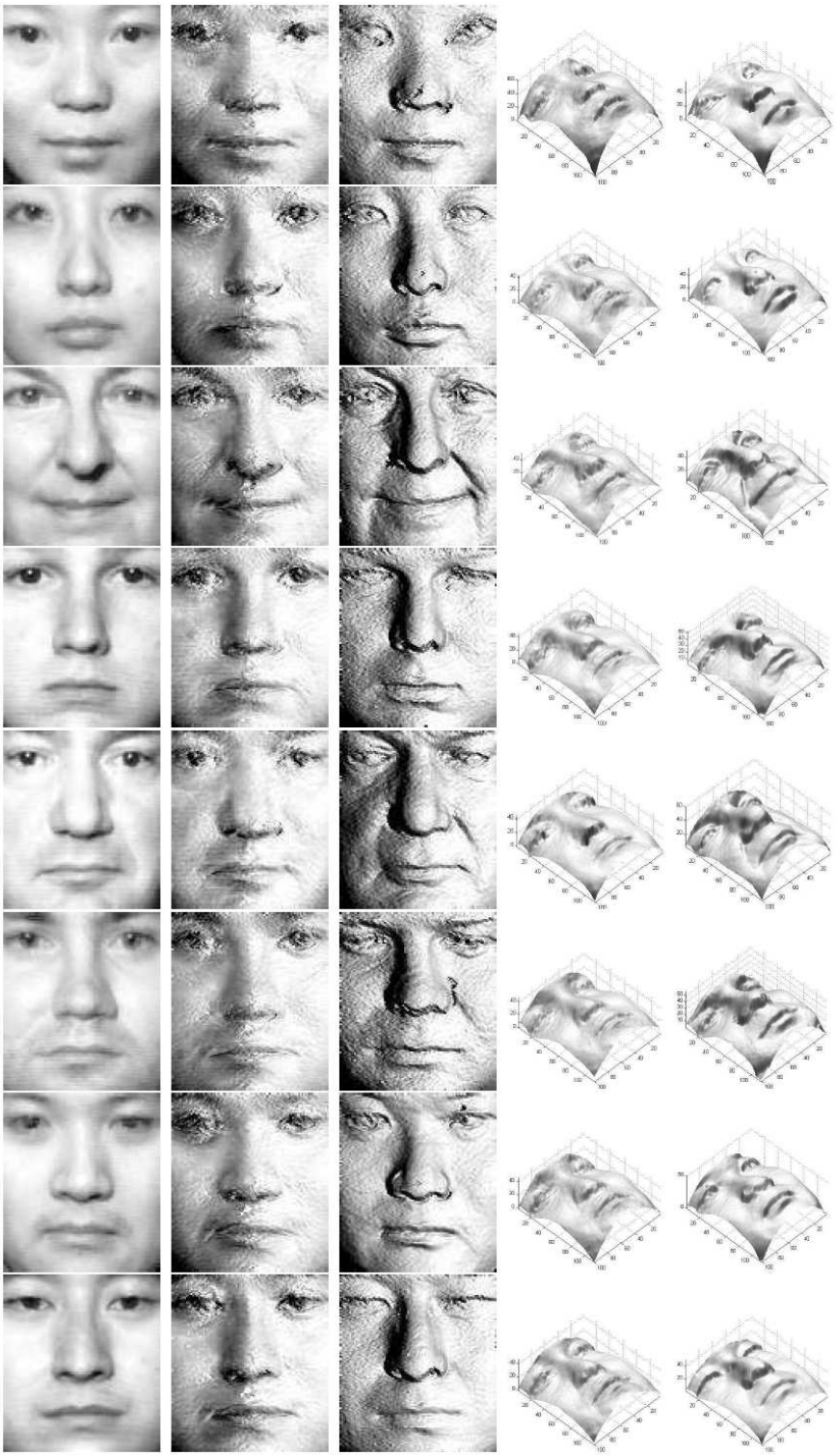

Figure 4.5: Examples of the results of SFS on Notre Dame images. From left to right are: the input intensity images, the recovered facial needle-maps, the ground-truth needle maps, the recovered surfaces, and the ground-truth surfaces [112].

range images to construct the statiscal model required in PGSFS.

Then, we apply the PGSFS method to the $2D$ images to recover facial needle-maps. The recovered needle-maps and integrated surfaces from eight facial images are shown in Figure 4.5.

4.5 Results

The aim of the experiments is to explore how the different distance measures can be used to measure different facets of facial shape including gender and expressions.

We compute Euclidean, cosine, Mahalanobis and geodesic distances for the long-vector representations of the needle-maps for the faces. We visualise the distribution of distances using Multi-Dimensional Scaling [54] to embed the faces in a two-dimensional pattern space (details of MDS were given in Section 4.2.1).

4.5.1 Gender Discrimination

In the last two decades, extensive research in gender difference systems has been conducted. We use a 2.5D representation of facial shape needle-maps which can be recovered from 2D facial images. Using the novel face representation to calculate the distance measures, this thesis gives a new contribution to the literature to recognise gender difference.

As explained before, we have two data sets. The first data set consists of 200 facial needle-maps extracted from range images in the Max Planck dataset (see Section 4.4). There are 100 females and 100 males, annotated with ground truth.

In Figures 4.6 corresponding Euclidean distance, Figure 4.7 corresponding cosine

distance, Figure 4.8 corresponding Geodesic distance and Figure 4.9 corresponding Mahalanobis distance. The first column shows the MDS embedding of the pattern of distances into a 2-dimensional space. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances. The results in the first row are obtained using Euclidean distance, the second row are for cosine distance, the third row for geodesic distance and the fourth row for Mahalanobis distance.

In Figures 4.6 corresponding Euclidean distance, Figure 4.7 corresponding cosine distance, Figure 4.8 corresponding Geodesic distance and Figure 4.9 corresponding Mahalanobis distance. The first column shows the MDS embedding of the pattern of distances into a 2-dimensional space. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances. The results in the first row are obtained using Euclidean distance, the second row are for cosine distance, the third row for geodesic distance and the fourth row for Mahalanobis distance.

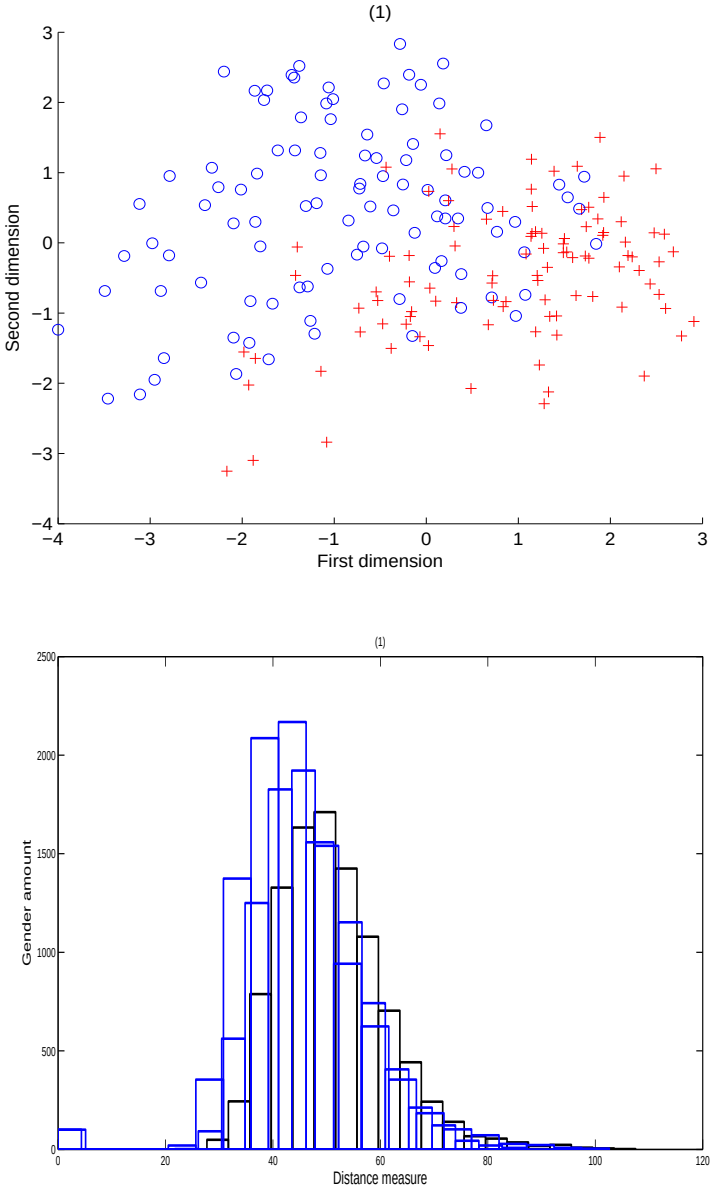

Figure 4.6: Max Planck database - Euclidean Distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

The second data set is EAR data set [36] [18] which comprises 200 (100 females and 100 males) laser scanned (Cyberware TM) human heads. The facial needle-maps are obtained by first orthographically projecting the facial range scans onto a frontal view plane, and then aligning the plane according to the eye centers, and cropping the plane to be 142-by-124 pixels to maintain only the inner part of the face. Finally, the surface normal at each pixel position is computed using the gradients of the processed range image. More informations, see Section 4.4.

In Figures 4.10 corresponding Euclidean distance, Figure 4.11 corresponding cosine distance, Figure 4.12 corresponding Geodesic distance and Figure 4.13 corresponding Mahalanobis distance. The first column shows the MDS embedding of the pattern of distances into a 2-dimensional space. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances. The results in the first row are obtained using Euclidean distance, the second row are for cosine distance, the third row for geodesic distance and the fourth row for Mahalanobis distance.

Turning our attention first to the embeddings (see Figures between 4.6 to 4.13), it is clear that in all four cases the distributions overlap, and gender discrimination can not be achieved using a simple class boundary. However, in the case of the geodesic distance, the distributions of male and female markers are concentrated in very different ways. In particular the female markers are more densely concentrated.

This would suggest that probabilistic separation may be feasible, and the unambiguous male subjects are separated from the female ones. It is worth noting that

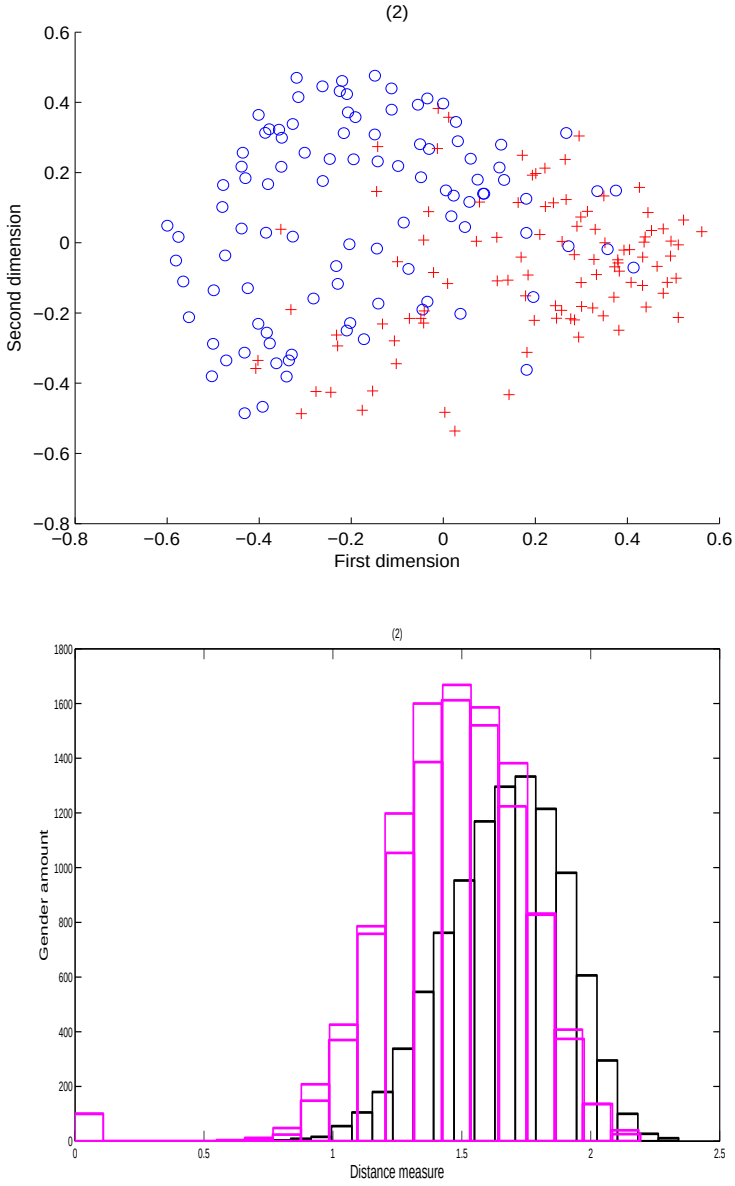

Figure 4.7: Max Planck database - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

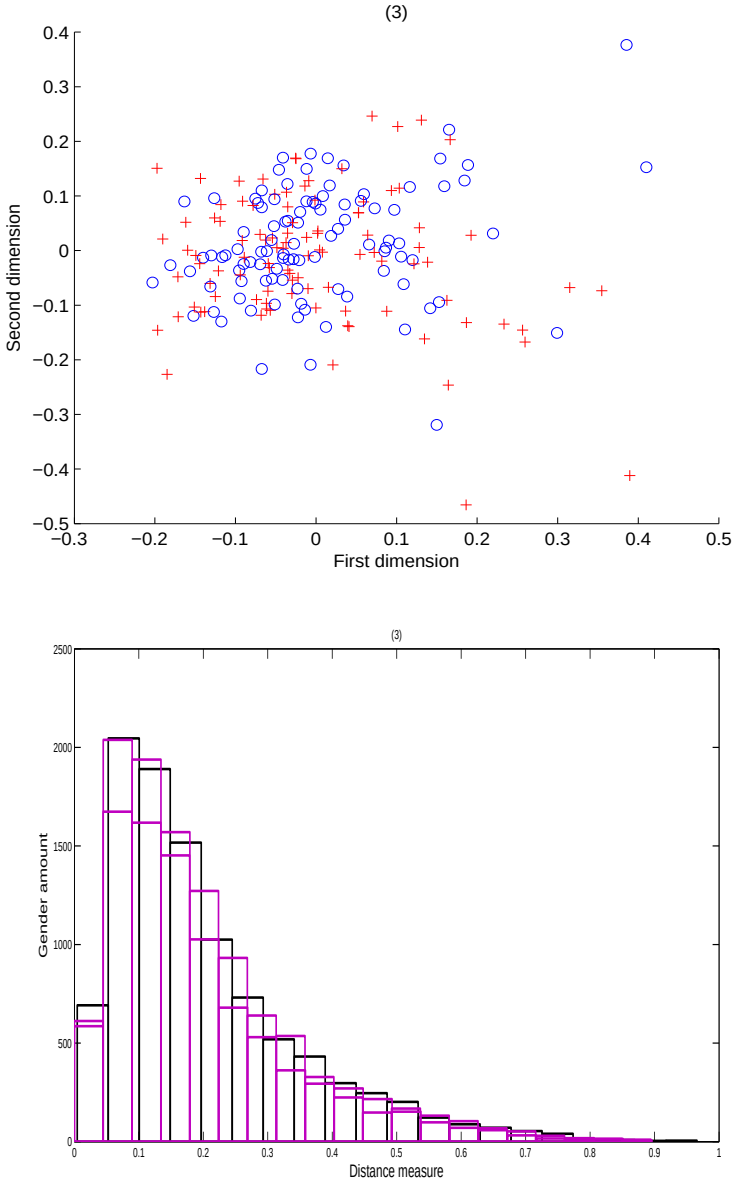

Figure 4.8: Max Planck database - geodesic distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

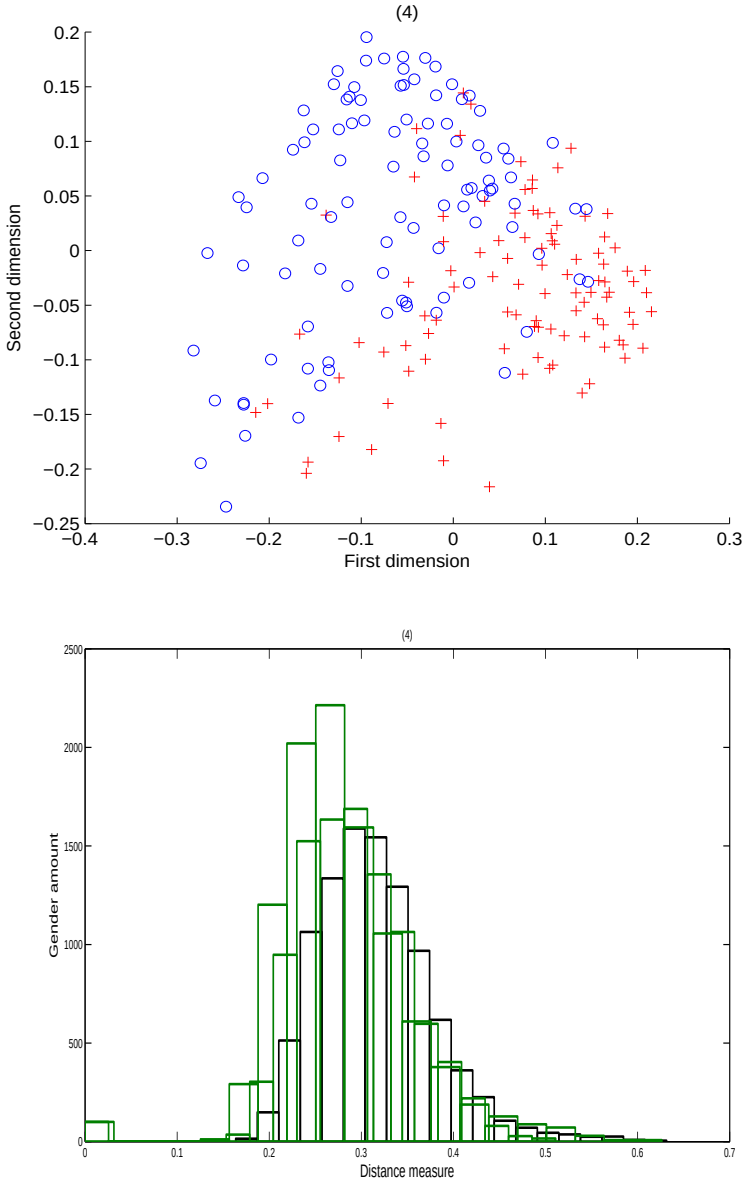

Figure 4.9: Max Planck database - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

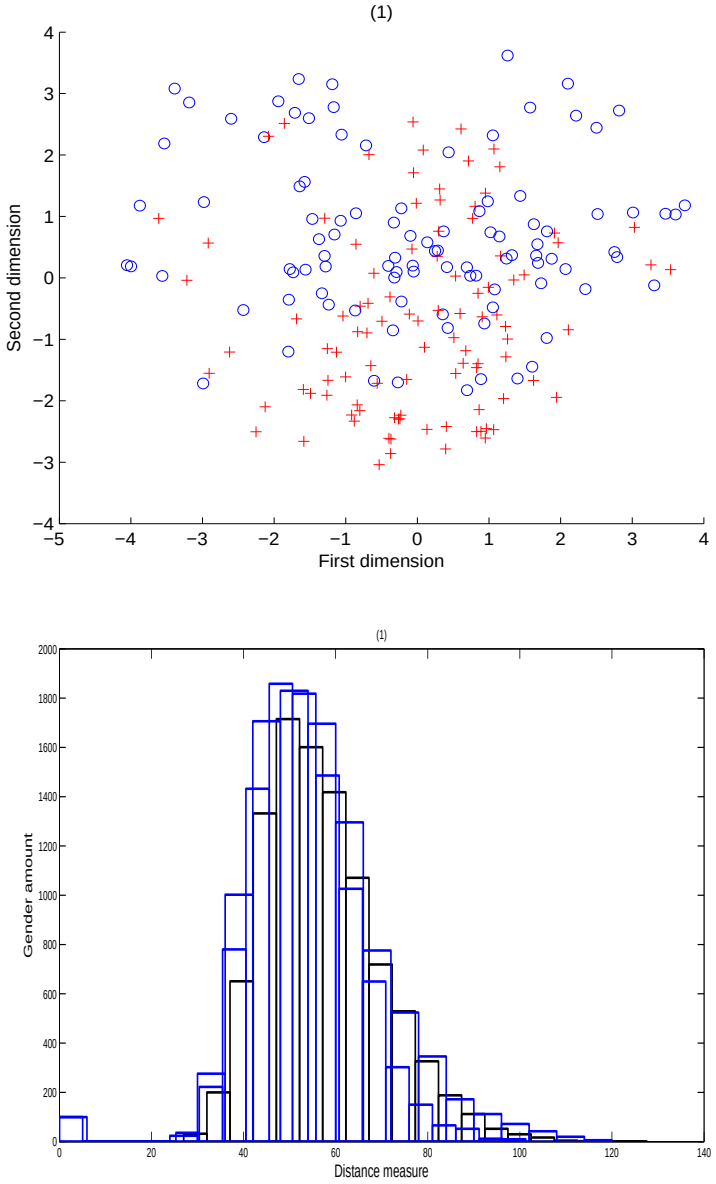

Figure 4.10: Notre Dame database - Euclidean distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

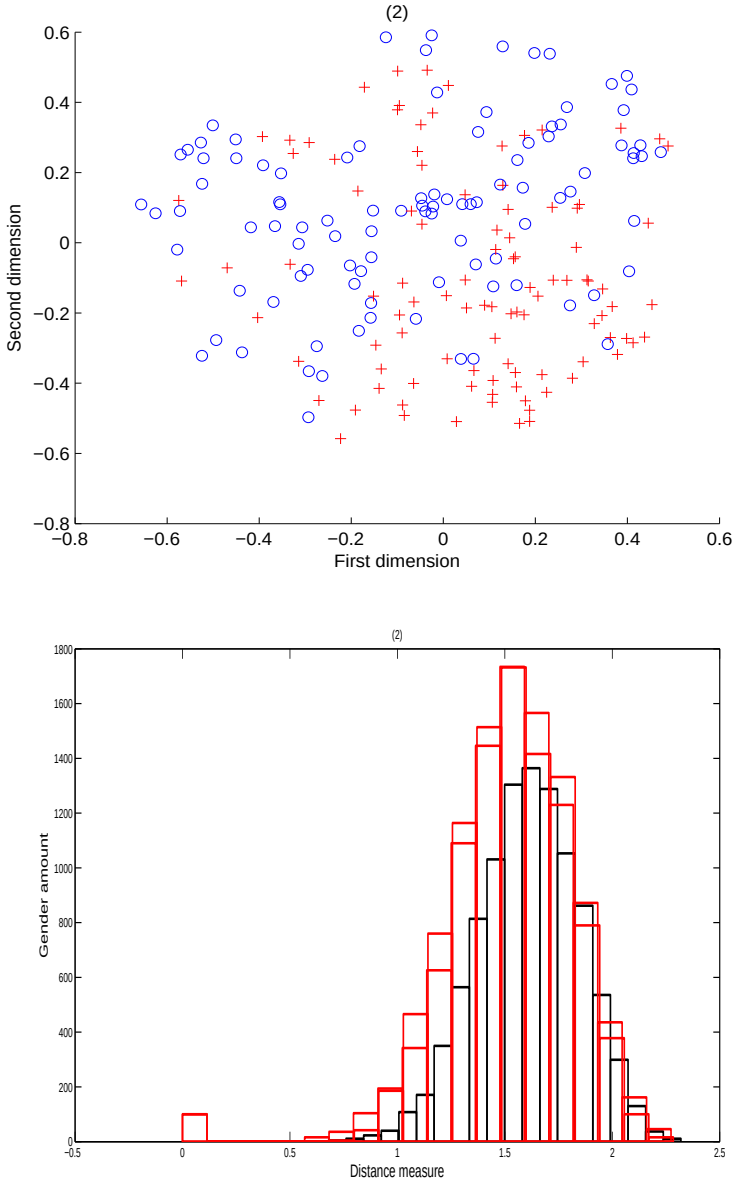

Figure 4.11: Notre Dame database - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

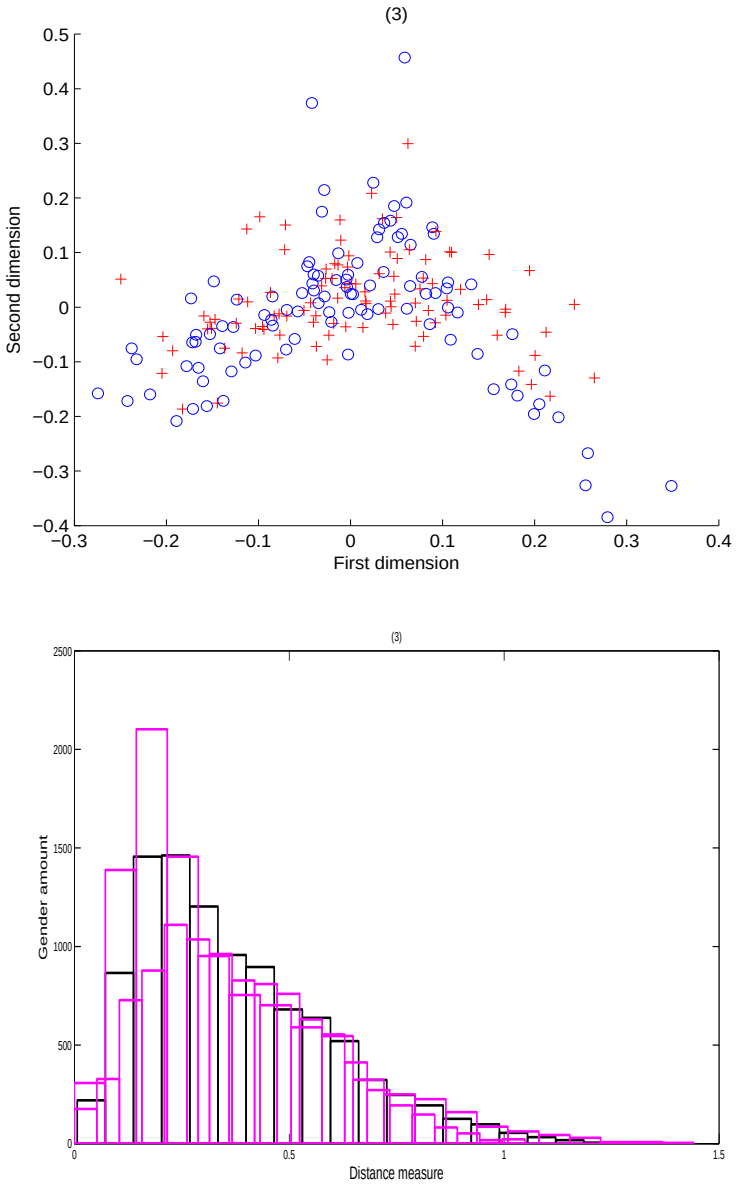

Figure 4.12: Notre Dame database - geodesic distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

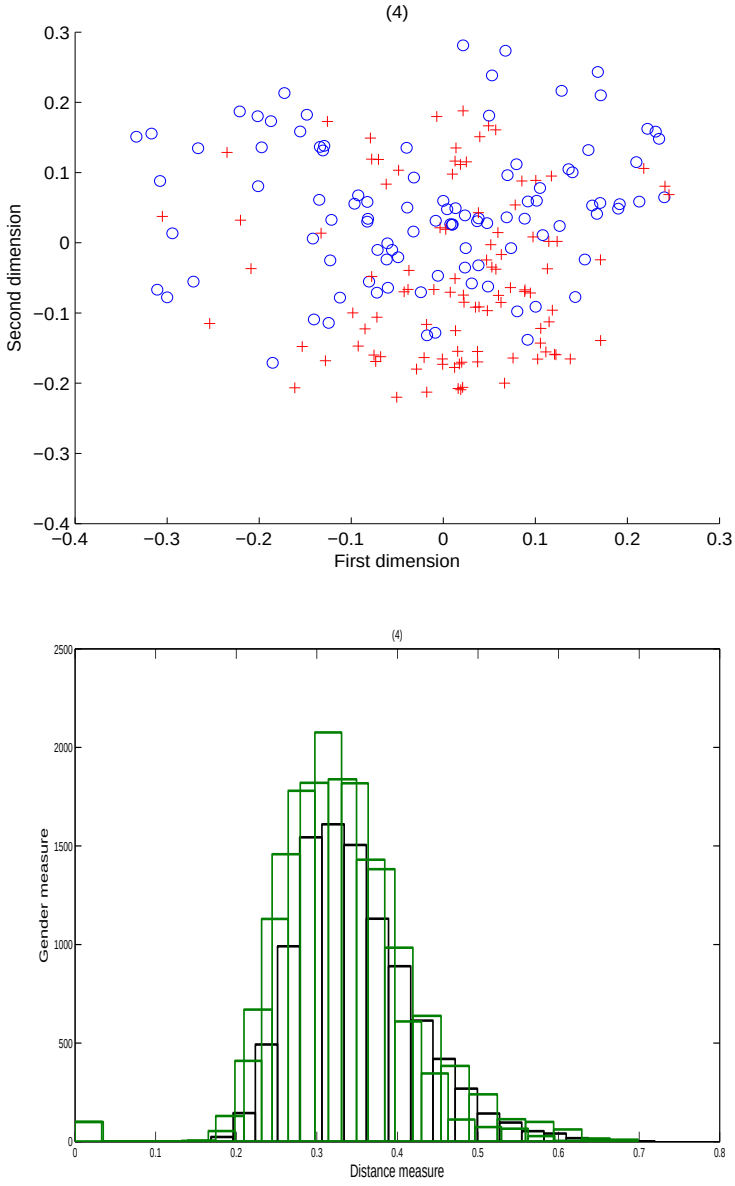

Figure 4.13: Notre Dame database - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of gender. The second column shows the Distance histograms. Gender - 100 Female and 100 Male. The blue markers are used to denote male subjects, and the red ones female subjects. The second column shows the corresponding distribution of distances as a histogram. The dark lines are between class distances, and the light lines are the distribution of within-class distances (male and female).

discriminating male and females faces on the basis of shape alone is a difficult task, and male human observers consider additional cues such as the hair-style.

These conclusions are borns by the distributions of distance. These are clearly overlapped, and only in the case of cosine and Mahalanobis distances is there a significant difference in mean within class and between class distance for the different genders.

4.5.2 Face expressions

The second experiment explores the ability of the distance measures to distinguish the same face when presented with a different expression. We use 5 different sets of data from the BU-3DFEDB database. Details about the database were described in Section 4.4. In this database, there are male and female subjects. We work with surface normals estimated from range-images.

The set of data (A) consists of 3 different faces each of which appears with 7 different facial expressions. The second data set (B) has 5 different faces with 7 different facial expressions. The data set (C) has 20 different faces with 7 different facial expressions. The data set (D) has 30 different faces with 7 different facial expressions and the final data set (E) has 40 different faces with 7 different facial expressions. In our MDS visualisations we show the different expressions for the same subject with the same symbol.

The results of our analysis are shown in Figures between 4.26 to 4.33 The first column shows the MDS embeddings of data-sets A to E and the second column shows the histogram between and within-class distance distributions for data-sets A

to E.

Figure 4.26 shows Euclidean distance, Figure 4.27 shows cosine distance, Figure 4.28 shows geodesic distance and Figure 4.29 shows Mahalanobis distance. Figures between 4.26 to 4.29 show the MDS embeddings and the histogram between and within-class distance distribution of data-set A. Each row represents a distance measure. First row shows the MDS embeddings and the second row shows the distances histogram.

Figure 4.18 shows Euclidean distance, Figure 4.19 shows cosine distance, Figure 4.20 shows geodesic distance and Figure 4.21 shows Mahalanobis distance. Figures between 4.18 to efef52 show the MDS embeddings and the histogram between and within class distance distribution of data-set B. Each row represents a distance measure. First row shows the MDS embeddings and the second row shows the distances histogram.

Figure 4.22 shows Euclidean distance, Figure 4.23 shows cosine distance, Figure 4.24 shows cosine distance and Figure 4.25 shows Mahalanobis distance. Figures between 4.22 to 4.25 show the MDS embeddings and the histogram between and within class distance distribution of data-set C. First row represents a distance measure and the second row represents the histograms.

Figure 4.26 shows Euclidean distance, Figure 4.27 shows cosine distance, Figure 4.28 shows cosine distance and Figure 4.29 shows Mahalanobis distance. Figures between 4.26 to 4.29 show the MDS embeddings and the histogram between and within class distance distribution of data-set C. First row represents a distance measure and the second row represents the histograms.

Figure 4.30 shows Euclidean distance, Figure 4.31 shows cosine distance, Figure

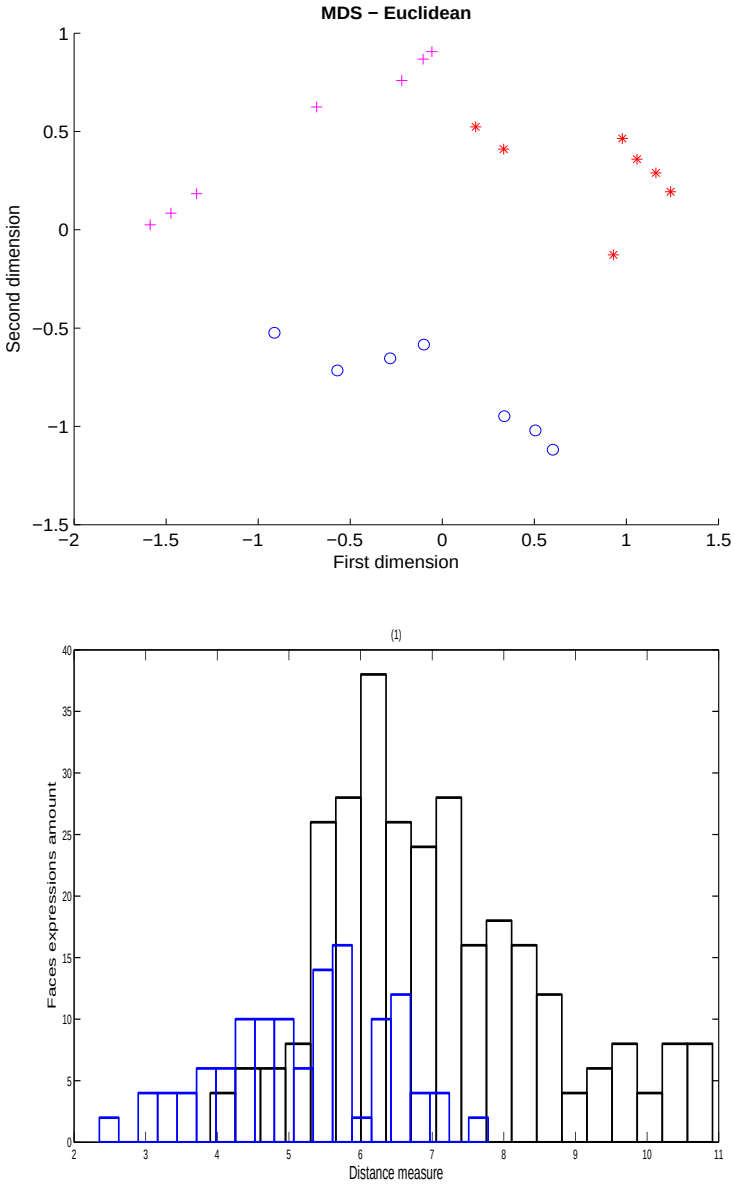

Figure 4.14: BU-3DFEDB database - Distance Measure - Euclidean distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of 3 faces and 7 expressions. The second column shows the Distance histograms.

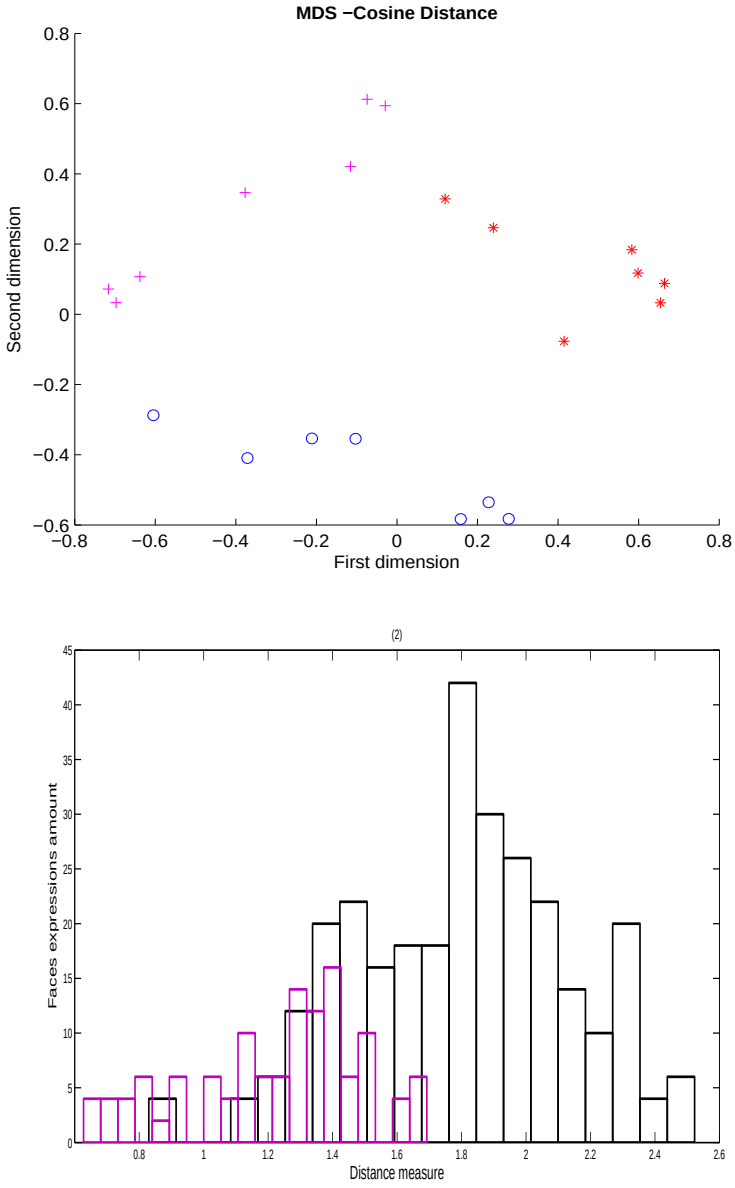

Figure 4.15: BU-3DFEDB database - Distance Measure - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 3 faces and 7 expressions. The second row shows the Distance histograms.

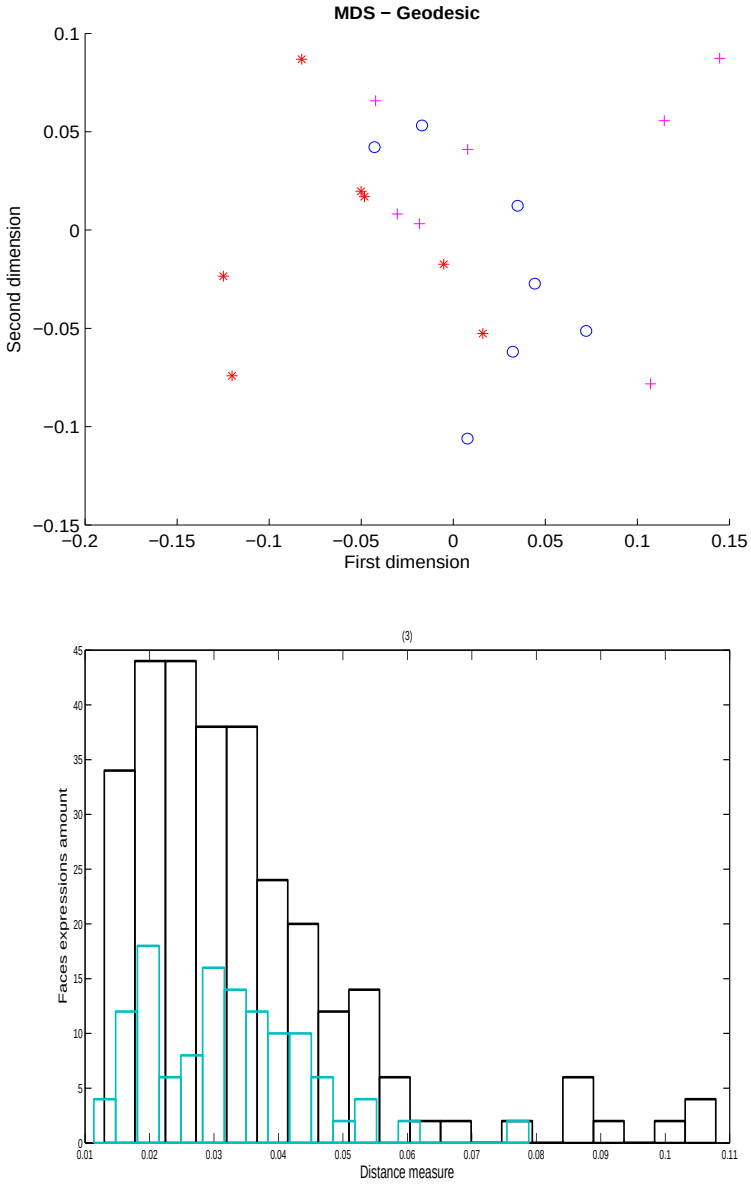

Figure 4.16: BU-3DFEDB database - Distance Measure - geodesic distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 3 faces and 7 expressions. The second row shows the Distance histograms.

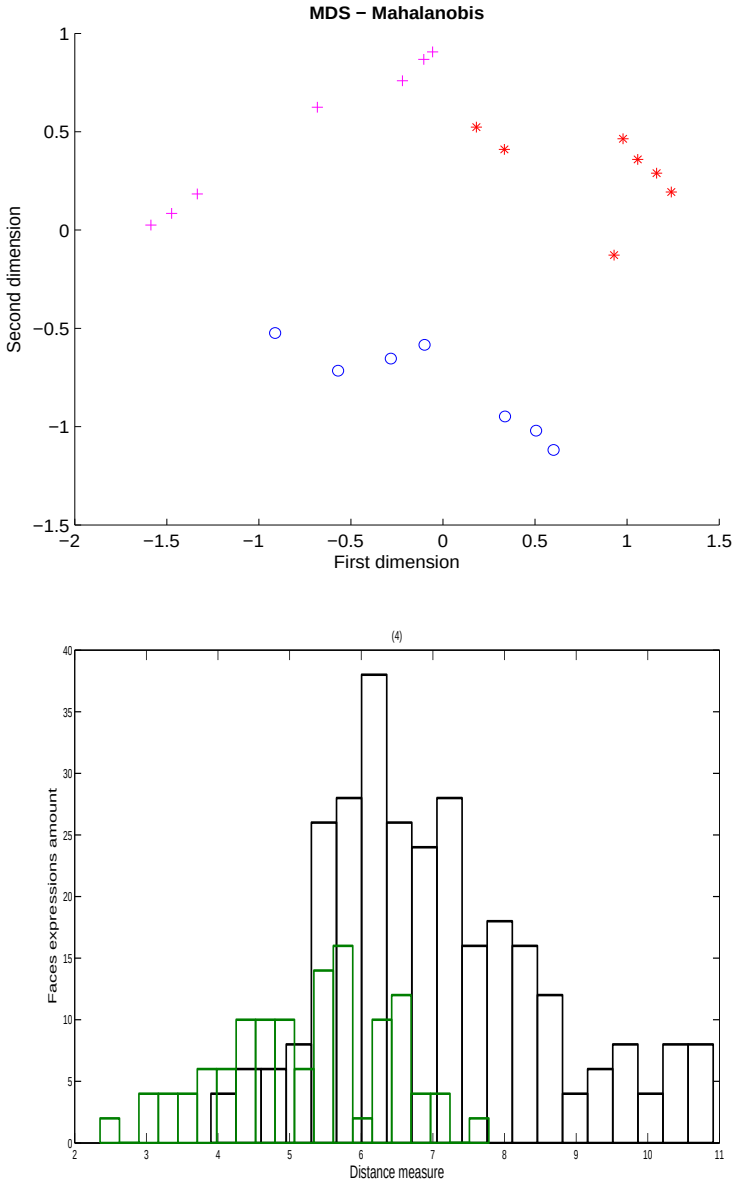

Figure 4.17: BU-3DFEDB database - Distance Measure - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 3 faces and 7 expressions. The second row shows the Distance histograms.

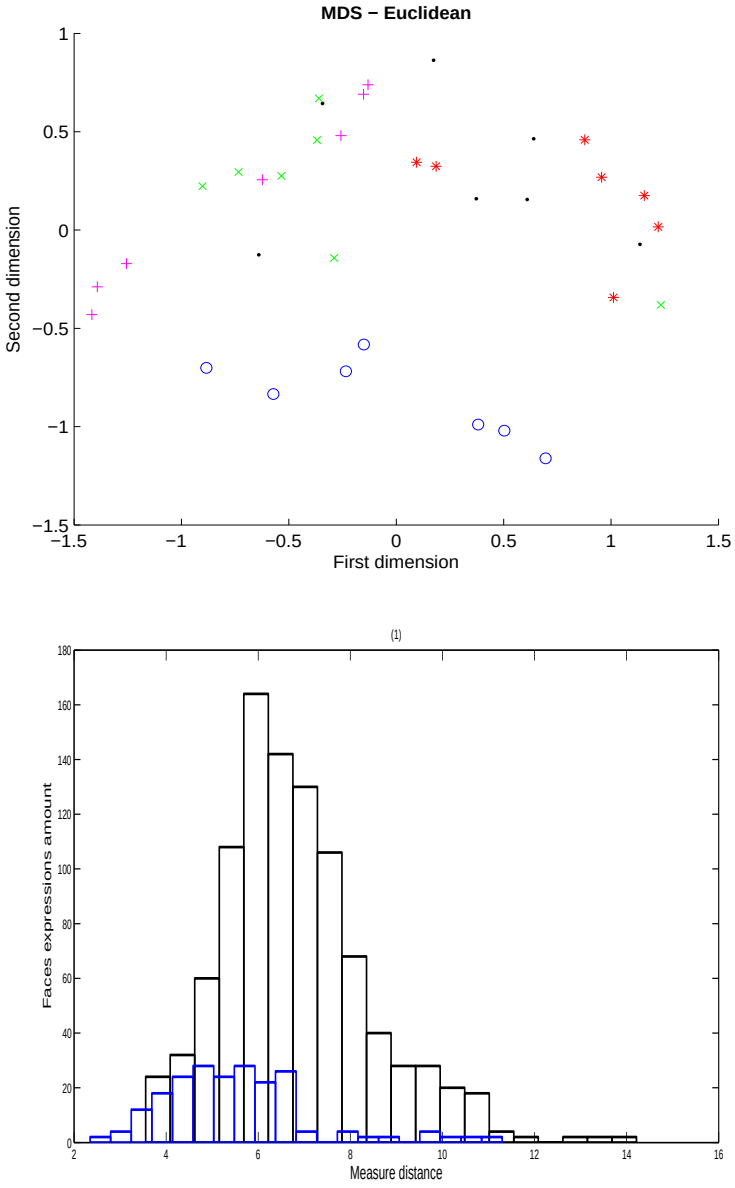

Figure 4.18: BU-3DFEDB database - Distance Measure - Euclidean distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 5 faces and 7 expressions. The second row shows the distance histograms.

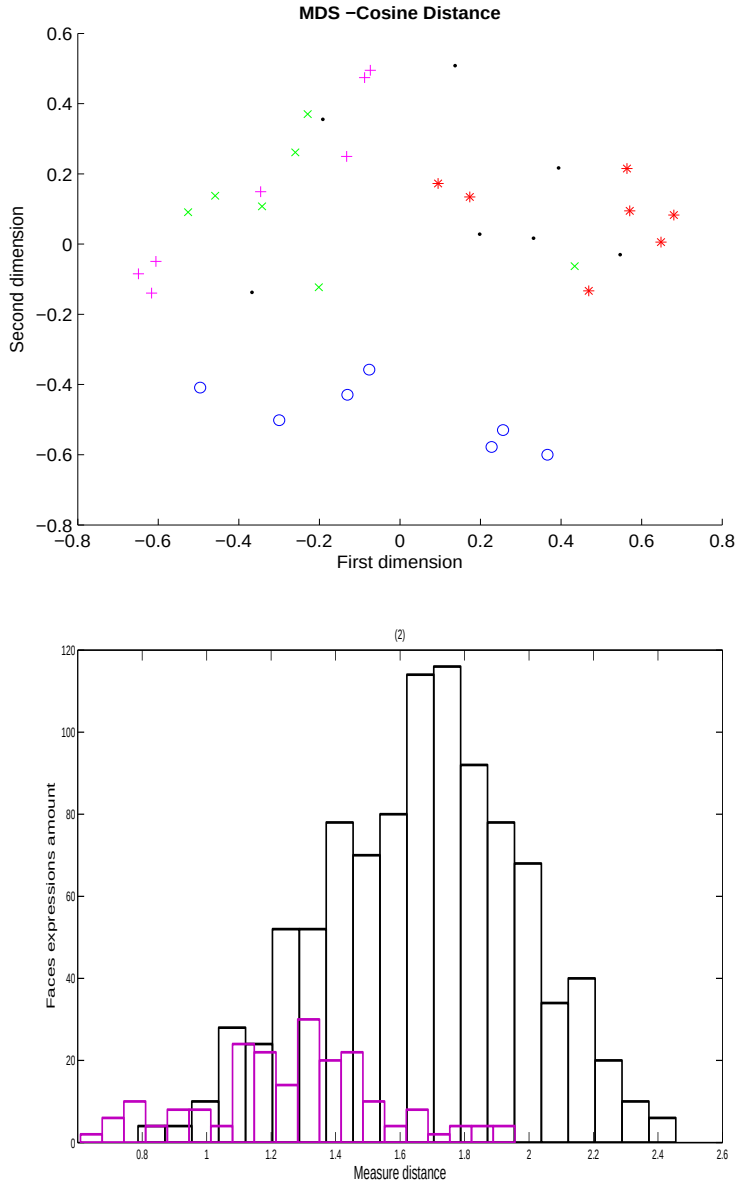

Figure 4.19: BU-3DFEDB database - Distance Measure - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 5 faces and 7 expressions. The second row shows the distance histograms.

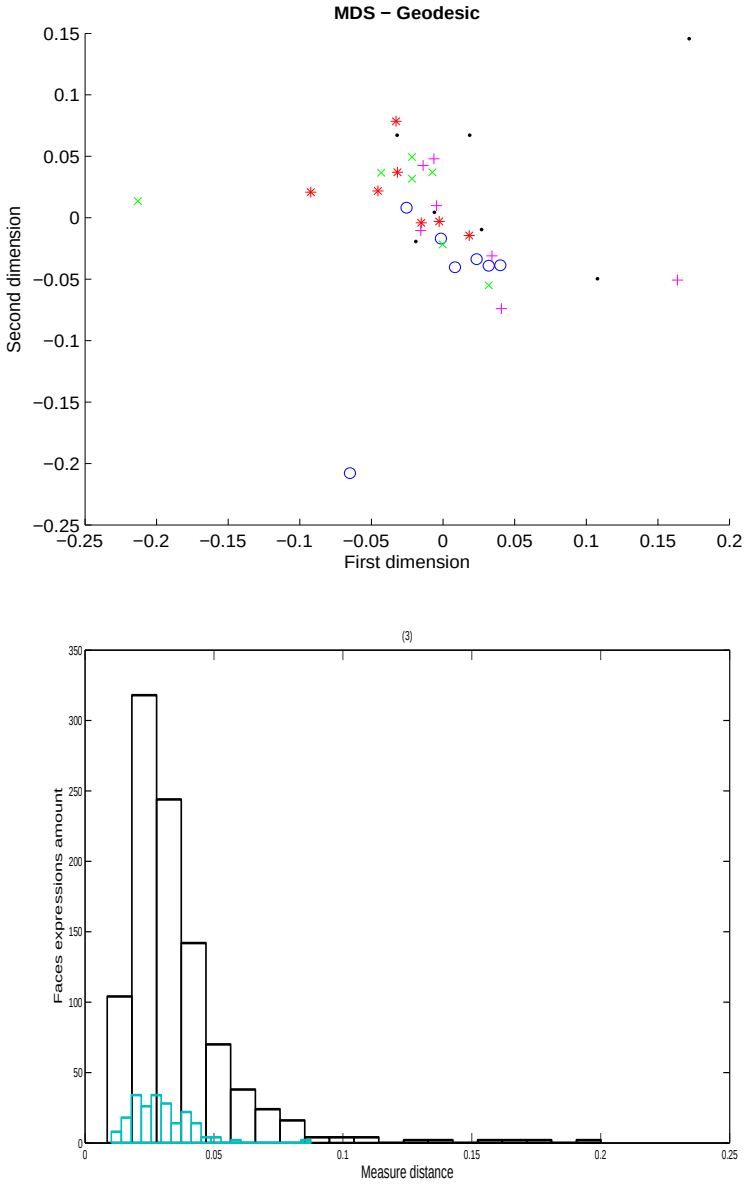

Figure 4.20: BU-3DFEDB database - Distance Measure - geodesic distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 5 faces and 7 expressions. The second row shows the distance histograms.

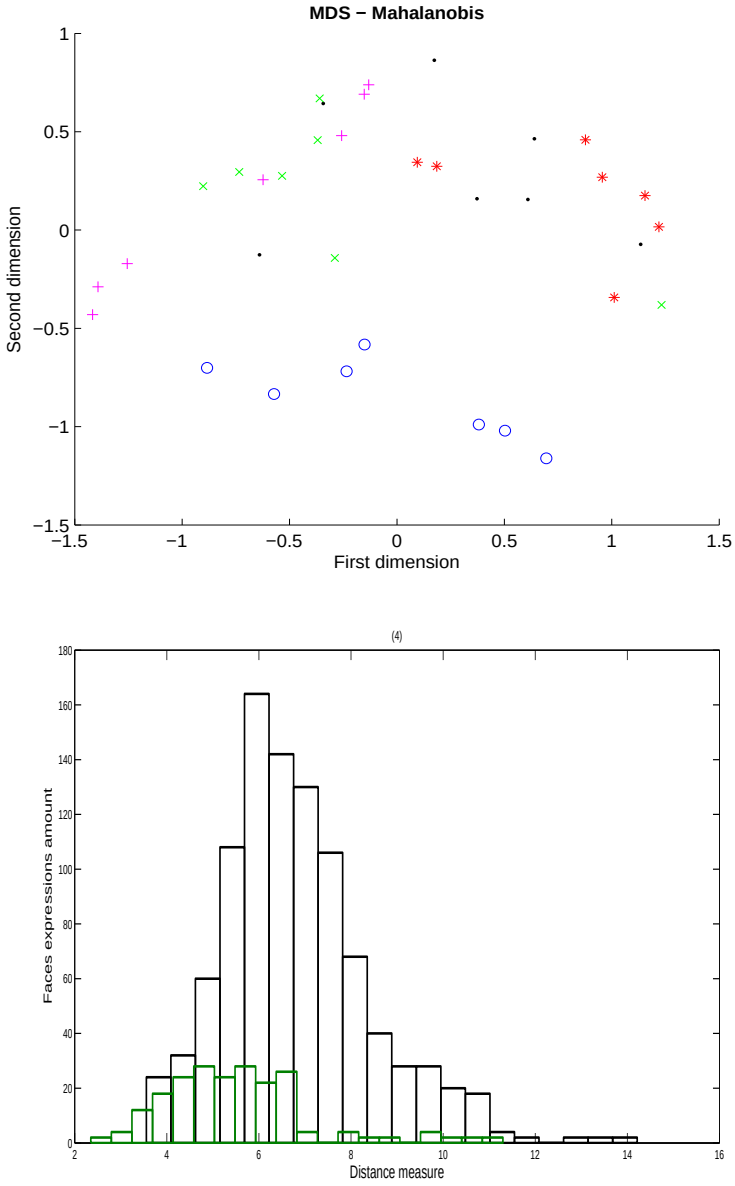

Figure 4.21: BU-3DFEDB database - Distance Measure - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 5 faces and 7 expressions. The second row shows the distance histograms.

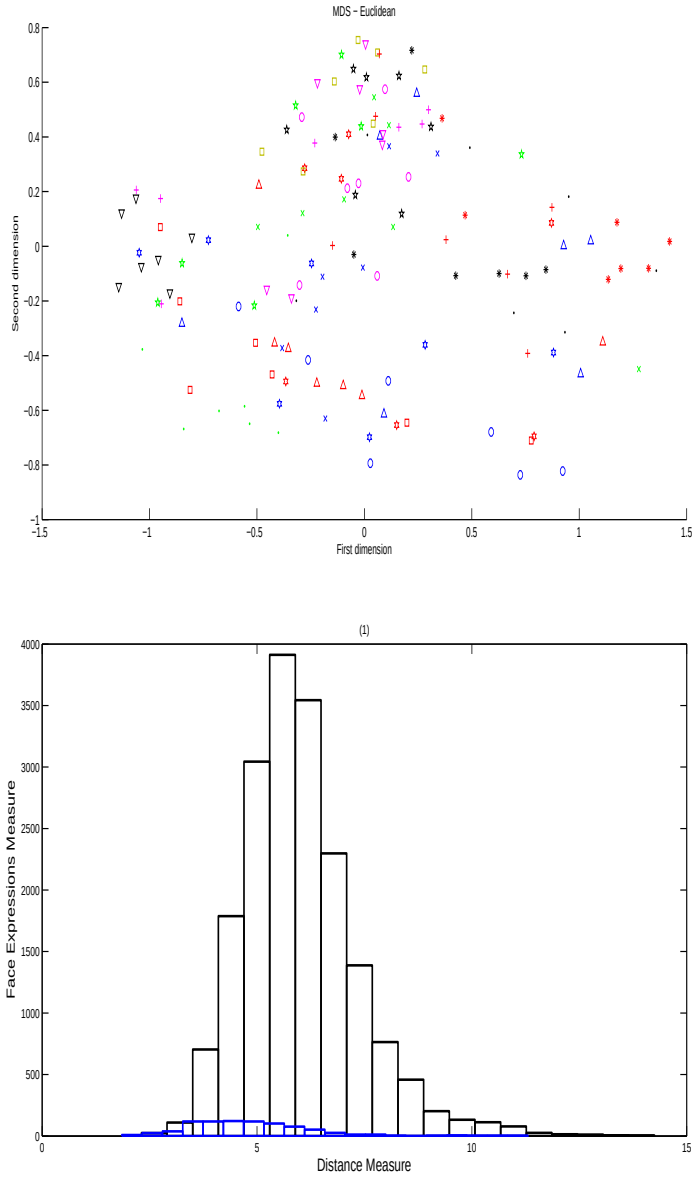

Figure 4.22: BU-3DFEDB database - Distance Measure - Euclidean distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 20 faces and 7 expressions. The second row shows the distance histograms.

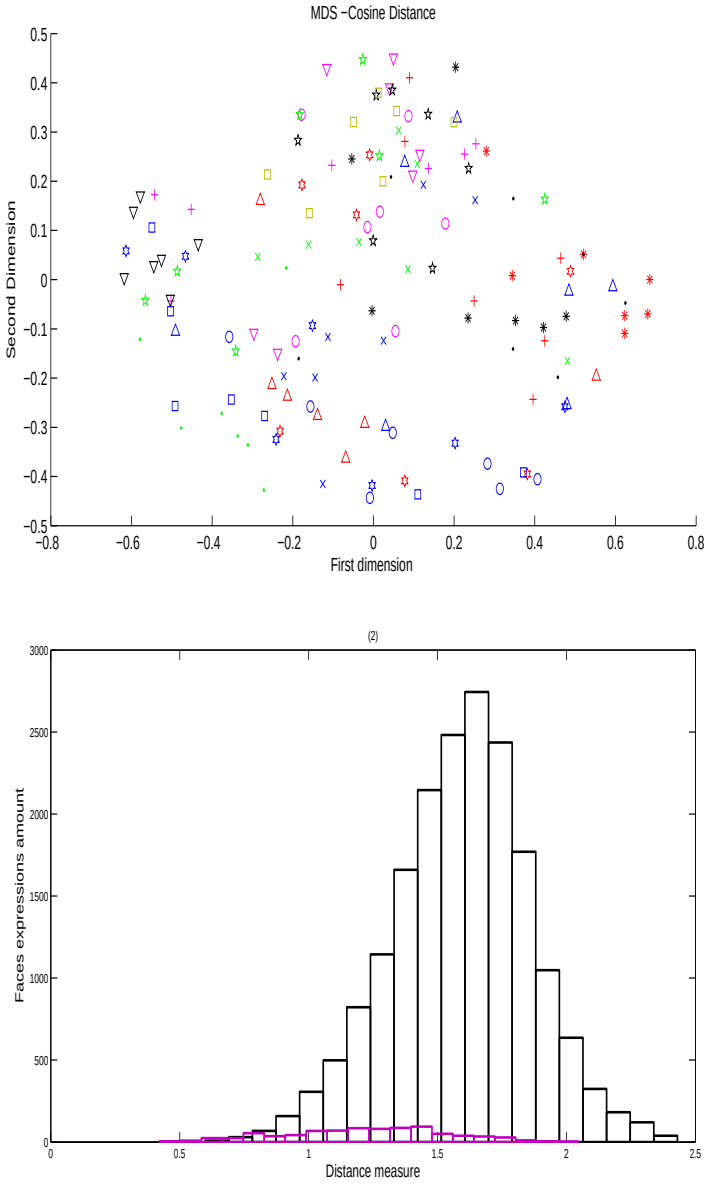

Figure 4.23: BU-3DFEDB database - Distance Measure - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 20 faces and 7 expressions. The second row shows the distance histograms.

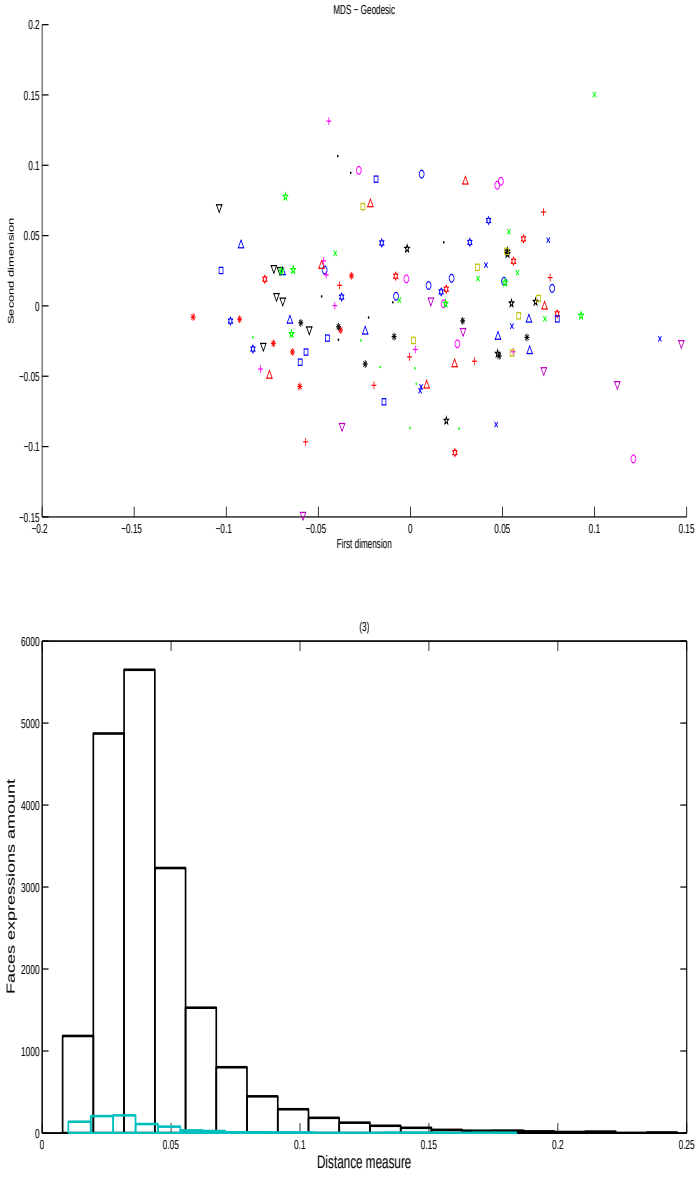

Figure 4.24: BU-3DFEDB database - Distance Measure - geodesic distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 20 faces and 7 expressions. The second row shows the distance histograms.

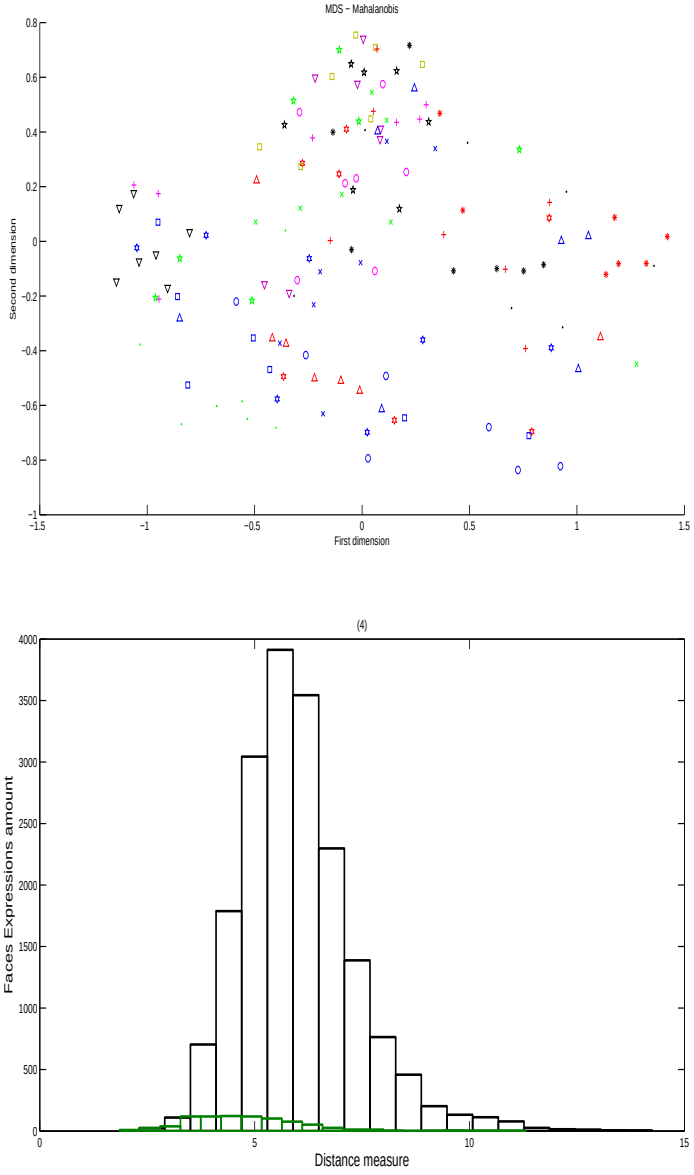

Figure 4.25: BU-3DFEDB database - Distance Measure - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 20 faces and 7 expressions. The second row shows the distance histograms.

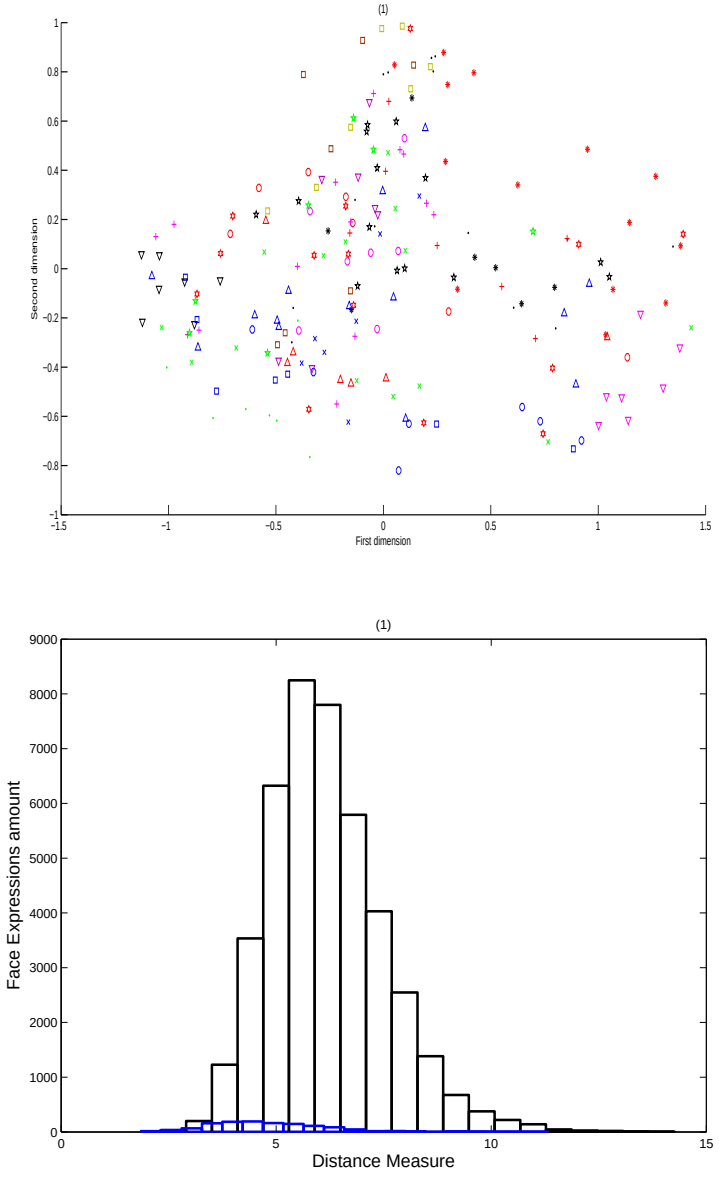

Figure 4.26: BU-3DFEDB database - Distance Measure - Euclidean distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 30 faces and 7 expressions. The second row shows the distance histograms.

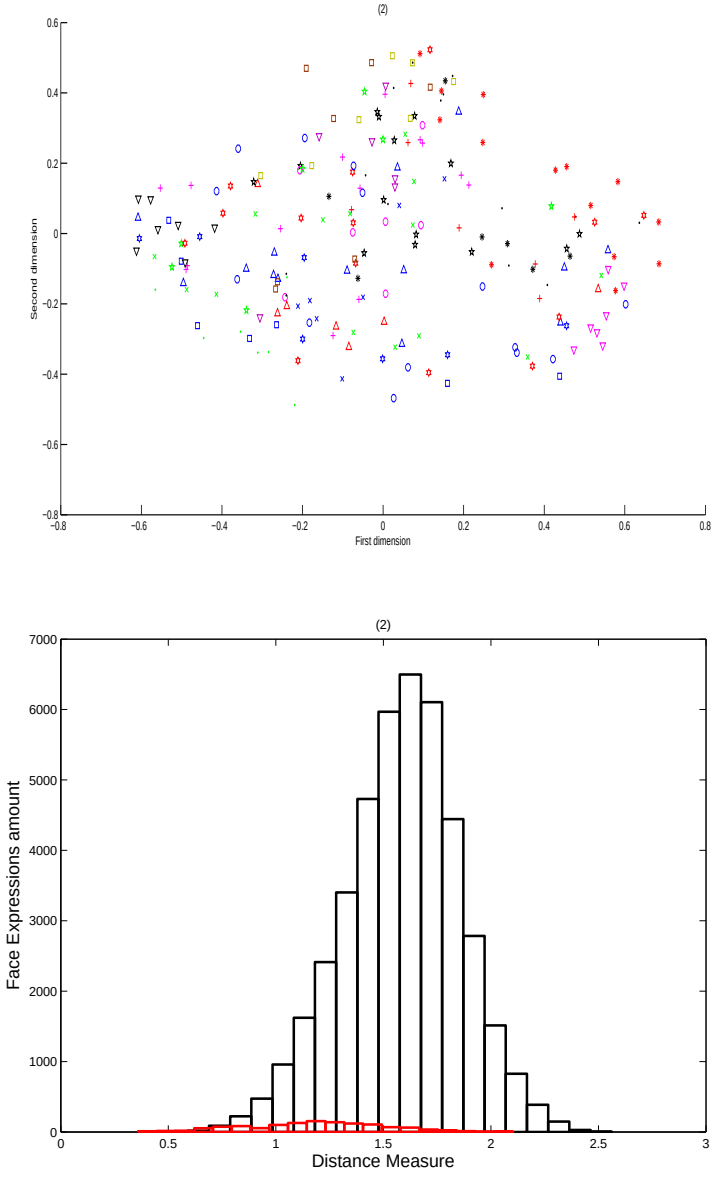

Figure 4.27: BU-3DFEDB database - Distance Measure - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 30 faces and 7 expressions. The second row the distance histograms.

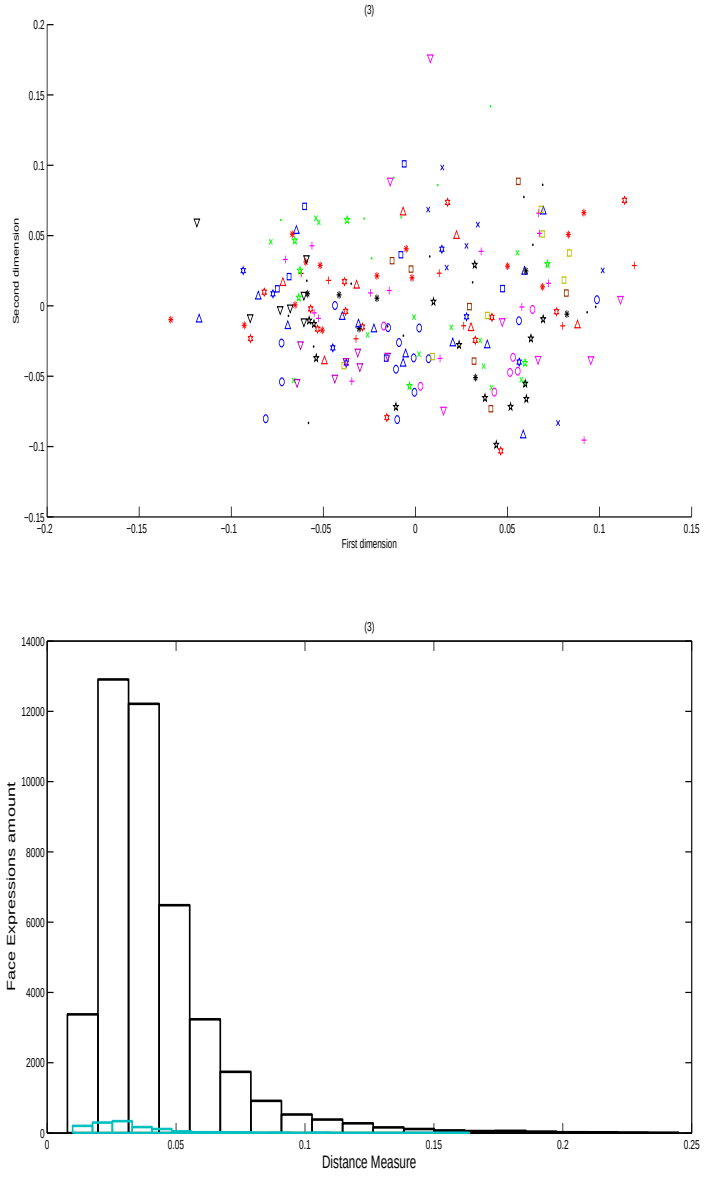

Figure 4.28: BU-3DFEDB database - Distance Measure - Geodesic Distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 30 faces and 7 expressions. The second row shows the Distance histograms.

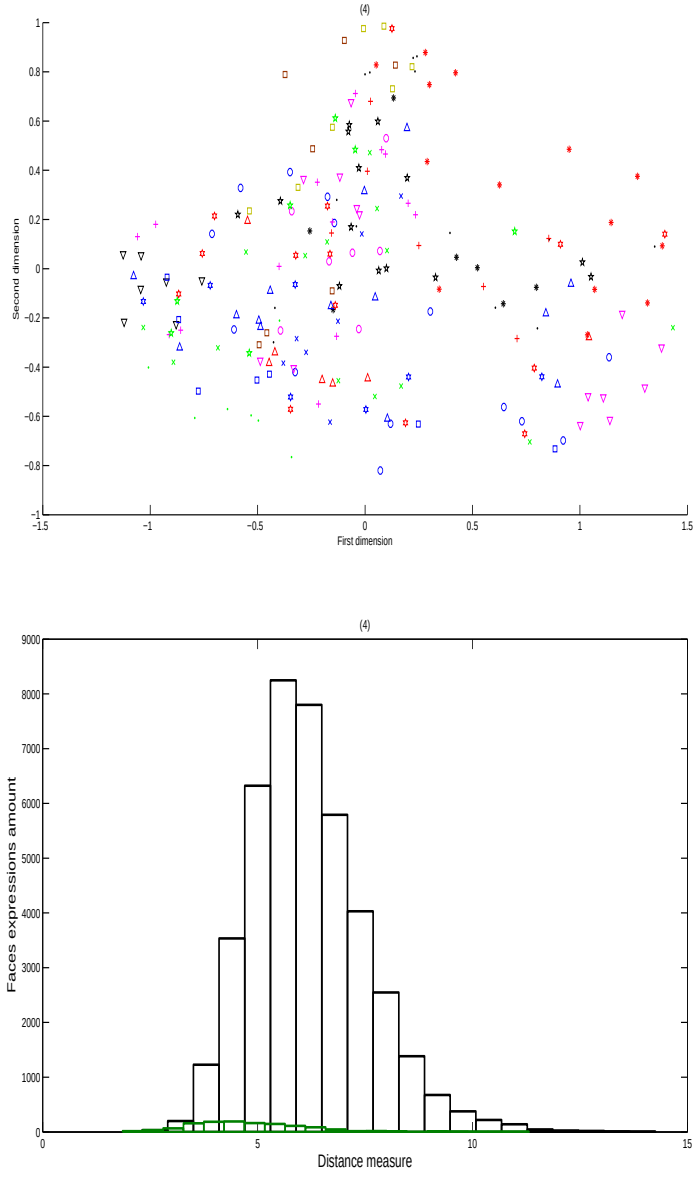

Figure 4.29: BU-3DFEDB database - Distance Measure - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 30 faces and 7 expressions. The second row shows the Distance histograms.

4.32 shows cosine distance and Figure 4.33 shows Mahalanobis distance. Figures between 4.30 to 4.33 show the MDS embeddings and the histogram between and within class distance distribution of data-set C. First row represents a distance measure and the second row represents the histograms.

Analyzing the MDS embeddings and the histograms of Figure 4.26, Figure 4.27, Figure 4.28 and Figure 4.29, we can see a good separation of the faces expressions for all the distances. Only, geodesic distance shows a low separation. In this case, the average for a small group of faces is satisfactory. Also, in Figure 4.30, Figure 4.31, Figure 4.32 and Figure 4.33, we visualize that all the distances (euclidean, cosine, geodesic and Mahalanobis) results start overlapping. Looking at the cosine histogram, the second row of Figure 4.31 gives the best separation of different faces under varying expression.

By increasing the number of faces (see Figures 4.22, 4.23, 4.24, 4.25, 4.26, 4.27, 4.28, 4.29, 4.30, 4.31, 4.32 and 4.33), the results are clearly unsatisfactory. Analyzing the histograms of those Figures (data set C: 20 faces and 7 expressions, data set D: 30 faces and 7 expressions and data set E: 40 faces and 7 expressions) we achieved a good result using cosine distance. On the other hand, euclidean, geodesic and Mahalanobis gives a similar result: almost indistinguishable.

An overview of the results, from the MDS embeddings, for a small number of faces data-set (see Figure 4.26, 4.27, 4.28 and Figure 4.29), we achieve a good separation of different faces under varying expression. In fact, the best separation is achieved with Mahalanobis distance. Also, Euclidean and cosine distances achieve good results. The face expression separation in those three cases are very satisfactory. In the other hand, geodesic distance shows an unsatisfactory result. However, as the number of

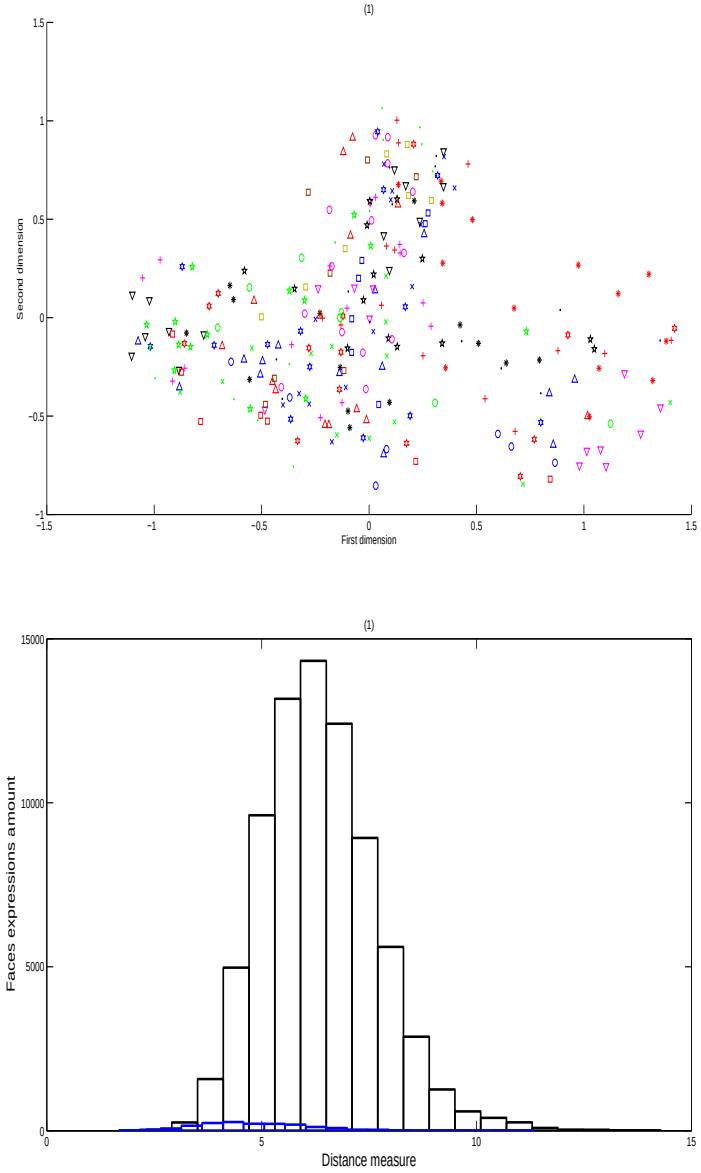

Figure 4.30: BU-3DFEDB database - Distance Measure - Euclidean distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of 40 faces and 7 expressions. The second column shows the distance histograms.

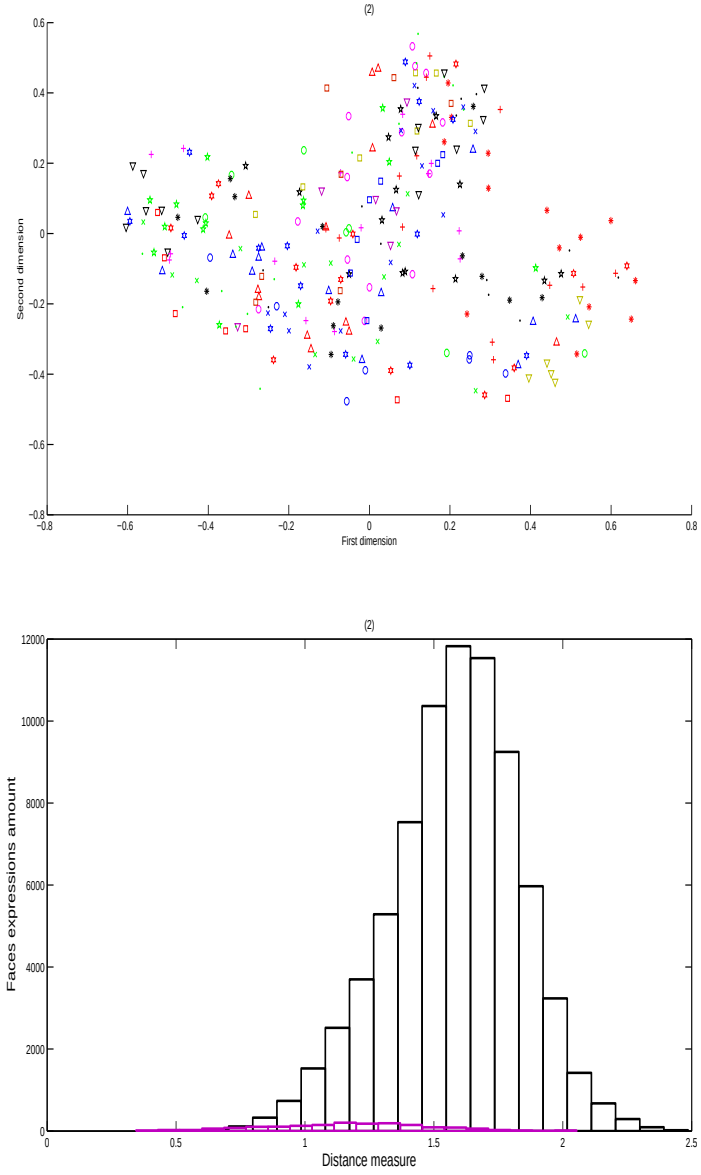

Figure 4.31: BU-3DFEDB database - Distance Measure - cosine distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first column shows the MDS analyis of 40 faces and 7 expressions. The second column shows the distance histograms.

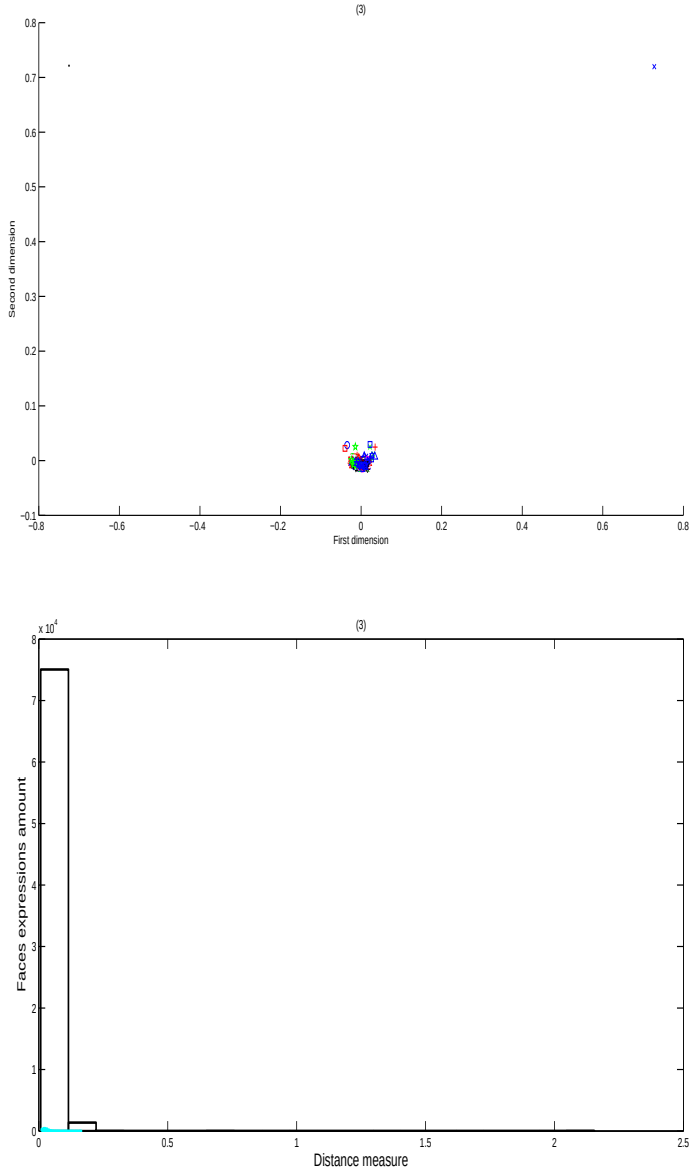

Figure 4.32: BU-3DFEDB database - Distance Measure - geodesic distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 40 faces and 7 expressions. The second row shows the distance histograms.

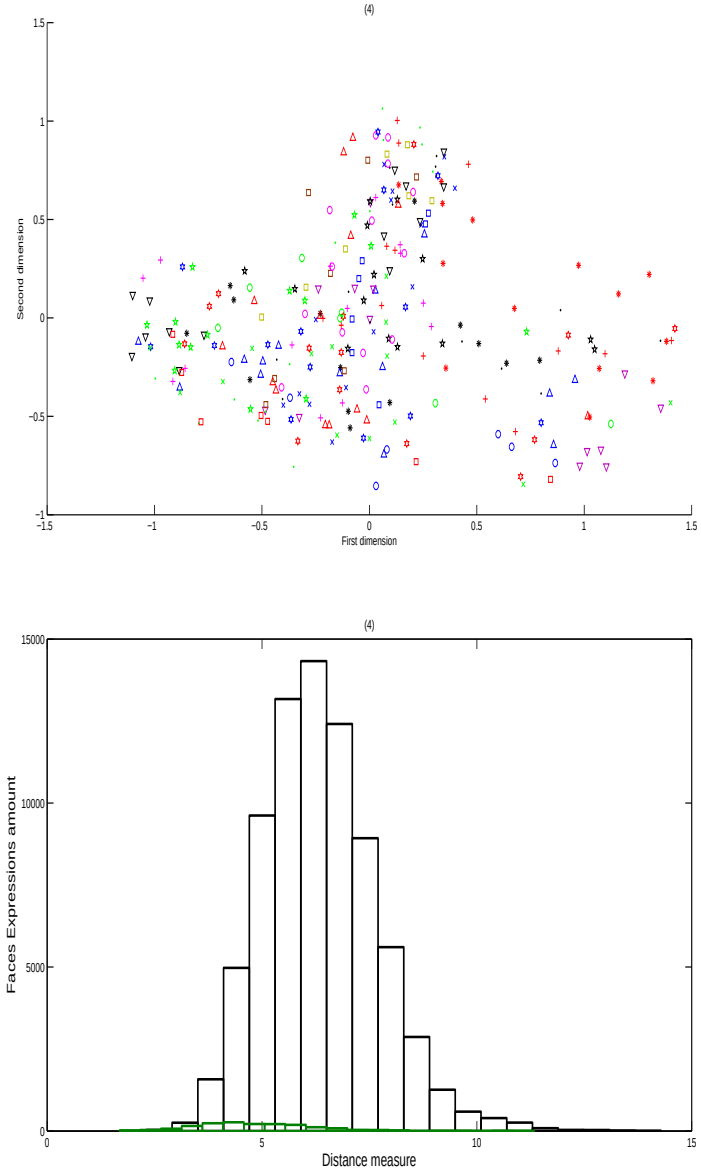

Figure 4.33: BU-3DFEDB database - Distance Measure - Mahalanobis distance. We use MDS to embed the pattern of distances in a 2-dimensional space. The first row shows the MDS analyis of 40 faces and 7 expressions. The second row shows the distance histograms.

faces increase, the overlap becomes significant (see Figures 4.22, 4.23, 4.24, 4.25, 4.26, 4.27, 4.28, 4.29, 4.30, 4.31, 4.32 and 4.33). It is clear that in the case of the Euclidean and cosine distances, while the distance distributions overlap, the mean interclass distance is less than the mean between class distance. In fact, the best separation is given by the cosine and Mahalanobis distances. For the geodesic distance this is not the case, and the distributions are almost indistinguishable.

4.6 Conclusions

In this chapter, we showed how to recognize face expressions and gender difference using 2.5D facial needle-maps. The needle-maps are extracted from intensity images using a model-based Shape-From-Shading method, referred to as Principal Geodesic Shape-From-Shading (PGSFS). The PGSFS method relies on a statistical model of facial shape formulated in the needle-map domain using Principal Geodesic Analysis. PGA is formulated using ideas from differential geometry and captures the non-Euclidean statistical nature of fields of unit-vectors. The PGSFS method is used to iteratively recover needle-maps that realistically capture facial shape and also satisfy the image irradiance equation as a hard constraint. Therefore, the recovered facial needle-maps both encode facial shape information and implicitly capture facial texture information.

For the recovery of facial shapes from 2D images, we have a special interest in statistical model based SFS and geometric SFS. SFS methods satisfying statistical model constraints [88] [89] [90] guarantee that the recovered shapes lie in a space representing realistic facial shapes, and therefore avoid invalid recoveries such as a

concaved nose. Geometric SFS methods [88] recover facial shapes satisfying data-closeness as a hard constraint. As a result, the recovered shapes implicitly encode the intensity information. This is especially useful for recognition tasks. Principal geodesic shape-from-shading [89] [90] method relies on a statistical model of facial shape formulated in the needle-map domain using principal geodesic analysis (PGA).

We demonstrate how to compute geodesic distance between fields of surface normals on a shape manifold and explore how the different distance measures can be used to measure different facets of facial shape including gender and expressions.We compute geodesic, Euclidean, Mahalanobis and cosine distances for the long-vector representations of the needle-maps of faces. We visualise the distribution of distances using MDS to embed the faces in a two-dimensional pattern space.

For gender experiments, Euclidean and Geodesic clearly overlapping, and only in the case of cosine and Mahalanobis distances is there a significant difference in mean within class and between class distance for the different genders.

Face Expression experiments show, for a small number of faces in data sets that, we achieve good separation of different faces under varying expression. In fact, the best separation is achieved with Mahalanobis distance. However, as the number of faces increases the overlapping becomes significant. It is clear that in the case of the Euclidean and cosine distances, while the distance distributions are overlapping, the mean interclass distance is less than the mean between class distance. In fact, the best separation is given by the cosine and Mahalanobis distances. For the geodesic distance this is not the case, and the distributions are almost indistinguishable.

This motivates our work in the following chapter. In Chapter 4, we focus on the development of statistical methods to recognize face expression and gender from

facial needle-maps.

CHAPTER 5

Fisher-Rao Metric to characterise facial shape

Applications in Computer Vision use Statistical Shape Analysis methods for the geometrical study of random objects where location, rotation and scale information can be removed. The subject is a new and exciting area of statistics, offering many fresh challenges ([24] [26] [91] [41]). The development of a rigorous statistical theory of shape began with the work by Kendall [49] which describes the shape formed by a set of random points under Brownian motion, and has been used in the statistical analysis of shape in archaeology and astronomy. Bookstein [10] and Ziezold [126], on the other hand, have developed methods for analysing the variations in biological forms. In the image analysis literature there are numerous examples of the use of Kendall's Shape Spaces [48], [87] and [102]. Recent developments in statistical

shape theory due to Small [86] and [87] suggest that improved Shape Spaces can be obtained by representing objects as points on a high-dimensional surface (a manifold) in such a way that different views of a given object correspond to a single point on the manifold.

Contributions

The aim in this chapter is to explore whether the Fisher-Rao metric can be used to measure different facets of facial shapes estimated from fields of surface normals using the von-Mises Fisher (vMF) distribution. The Fisher information of the family defines on the manifold a Riemannian metric known as the Fisher-Rao metric. Fisher-Rao has an approximation metric which is accurate if the variance of the measured errors is small. It is shown that the manifold of parameter values has a finite volume under the approximating metric.

In particular we aim to characterise the shape changes due to differences in gender and facial expressions. We make use of the vMF distribution since we are dealing with surface normal data over the sphere $\Re^2$. We choose to work with the von-Mises Fisher distribution because it is the natural probability distribution for high-dimensional directional data. The space of vMF distributions forms a differentiable manifold, which can be considered to be embedded in a higher dimensional space [93]. The embedding space induces a metric on the manifold that allows for an intrinsic way to measure distances on the manifold. A Riemannian manifold is a smooth manifold supplied with a Riemannian metric [45],[73].

The remainder of this chapter is organized as follows: Section 5.1 describes the Von-Mises Fisher distribution (vMF). Section 5.2 describes the Fisher Information

Matrix method. Next, Section 5.3 is concerned with assessing shape variation in fields of surface normals due to both facial expression and gender difference. In the case of facial expression we aim to explore the changes in facial shape due to subjects pulling seven different expressions namely, happiness, sadness, surprise, fear, anger, disgust and neutral. We also aim to explore if the techniques outlined in this work can be used to distinguish the gender of different subjects. Finally, Section 5.5 concludes this chapter with a summary.

5.1 The von-Mises Fisher distribution (vMF)

A d-dimensional unit random vector x (i.e., x) is said to have variate von Mises-Fisher (vMF) distribution if its probability density function is given by:

$$f_p(\underline{x}, \underline{\mu}, \kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \exp(\kappa \underline{\mu}^T \underline{x}). \tag{5.1}$$

where $\underline{x}$ is a p dimensional vector residing on the hyper-sphere S^{p-1} submersed in $\Re^p$. $I_l(\kappa)$ is the modified Bessel function of the first kind of order l. The density $f(x|\underline{\mu}, \kappa)$ is parameterized by the mean direction $\underline{\mu}$, and the concentration parameter κ, so called because it characterizes how strongly the unit vectors draw according to $f(x|\underline{\mu}, \kappa)$ are concentrated about the mean direction μ. Larger values of κ imply stronger concentration of the mean direction. In particular when $\kappa = 0$, $f(x|\underline{\mu}, \kappa)$ reduces to the uniform density on S^{d-1}, and as $\kappa \to \infty$, $f(x|\underline{\mu}, \kappa)$ tends to a point density. The interested reader is referred to Mardia et.al (2000) [61] for details on vMF distributions. The distribution is unimodal and rotationally symmetric around

the direction $\underline{\mu}$. Finally, the distribution is uniform over the hyper-sphere for $\kappa = 0$. The maximum likelihood estimators for the two parameters are obtained as follows. Suppose we have m samples of $\underline{x}$, i.e., $\underline{x}_1, ...\underline{x}_m$. The estimator of the mean direction is given by $\underline{\mu} = \frac{\sum_{i=1}^{m} \underline{x}_i}{||\sum_{i=1}^{m} \underline{x}_i||}$

There is no closed form estimator for the concentration parameter $\hat{\kappa}$. Instead, it is the solution of the transcendental equation

$$\frac{I_{\frac{p}{2}}(\hat{\kappa})}{I_{\frac{p}{2}-1}(\hat{\kappa})} = \frac{1}{m}||\sum_{i=1}^{m} \underline{x}_i||$$

In practice we solve this equation using the Newton-Raphson method [28]. It is worth noting that Jupp and Mardia [61] have developed some non-iterative approximations which apply under small and large values of κ. For p=3, the distribution is referred to as the vMF distribution.

5.2 Fisher Information Matrix

The Fisher information matrix is a Riemannian metric which can be defined on a smooth statistical manifold, i.e., a smooth manifold whose points are probability measures defined on a common probability space [64],[75],[66].

Let I = [0,1] and p: I $\times$ $\Re^k$ $\rightarrow$ $\Re^+$, (x,$\underline{\theta}$) $\mapsto$ p(x;$\underline{\theta}$), a k-dimensional family of positive probability density functions parameterized by the vector of parameters $\underline{\theta} = (\theta_1, ..., \theta_k)^T \in \Re^k$. In classical information geometry the Riemannian structure of the parameter space $\Re^k$ is defined by the Fisher information matrix with elements

$$g_{ij}(\underline{\theta}) = \int p(\underline{x}|\underline{\theta}) \frac{\partial}{\partial \theta_i} \log p(\underline{x}|\underline{\theta}) \frac{\partial}{\partial \theta_j} \log p(\underline{x}|\underline{\theta}) d\underline{x}. \tag{5.2}$$

The notation ∂_{θ_i} is used for the partial derivative with respect to the component θ_i of $\underline{\theta}$, where $\underline{\theta}$ is a vector of parameters associated with the density p. The Fisher-Rao metric tensor (5.2) is an intrinsic measure, allowing us to analyze a finite, k-dimensional statistical manifold M without considering how M resides in an R^{2k+1} space. In our case, we have 4 parameters and $\underline{\theta} = (\kappa, \mu_1, \mu_2, \mu_3)^T$, where $\mu = (\mu_1, \mu_2, \mu_3)^T$ is the density parameter vector.

In practice we divide each field of surface normals into windows whose size is determined by the overall image size. In our experiments, the window size is 4x4. This provides sufficient statistics to make stable estimates of the mean direction and concentration parameter. Fisher-Matrix g_{ij} is a matrix 4×4. We are using von mises with order the parameters as:

$$\begin{bmatrix} \theta^1 \\ \theta^2 \\ \theta^3 \\ \theta^4 \end{bmatrix} = \begin{bmatrix} \kappa \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

Then, the remaining entries are $4X4$ matrix are obtained by symmetric. We have to compute the integral for each entry.

$$g_{ij}(\theta) = \int p(X|\theta) \frac{\partial}{\partial \theta^i} \log p(X|\theta) \frac{\partial}{\partial \theta^j} \log p(X|\theta) dx \tag{5.3}$$

$$g_{1,1} = \int p(X|\theta) \frac{\partial}{\partial \kappa} \log p(X|\theta) \frac{\partial}{\partial \kappa} \log p(X|\theta) dx \tag{5.4}$$

$$g_{1,2} = \int p(X|\theta)\frac{\partial}{\partial\kappa}\log p(X|\theta)\frac{\partial}{\partial\mu_1}\log p(X|\theta)dx \tag{5.5}$$

$$g_{1,3} = \int p(X|\theta)\frac{\partial}{\partial\kappa}\log p(X|\theta)\frac{\partial}{\partial\mu_2}\log p(X|\theta)dx \tag{5.6}$$

$$g_{1,4} = \int p(X|\theta)\frac{\partial}{\partial\kappa}\log p(X|\theta)\frac{\partial}{\partial\mu_3}\log p(X|\theta)dx \tag{5.7}$$

$$g_{2,2} = \int p(X|\theta)\frac{\partial}{\partial\mu_1}\log p(X|\theta)\frac{\partial}{\partial\mu_1}\log p(X|\theta)dx \tag{5.8}$$

$$g_{2,3} = \int p(X|\theta)\frac{\partial}{\partial\mu_1}\log p(X|\theta)\frac{\partial}{\partial\mu_2}\log p(X|\theta)dx \tag{5.9}$$

$$g_{2,4} = \int p(X|\theta)\frac{\partial}{\partial\mu_1}\log p(X|\theta)\frac{\partial}{\partial\mu_3}\log p(X|\theta)dx \tag{5.10}$$

$$g_{3,3} = \int p(X|\theta)\frac{\partial}{\partial\mu_2}\log p(X|\theta)\frac{\partial}{\partial\mu_2}\log p(X|\theta)dx \tag{5.11}$$

$$g_{3,4} = \int p(X|\theta)\frac{\partial}{\partial\mu_2}\log p(X|\theta)\frac{\partial}{\partial\mu_3}\log p(X|\theta)dx \tag{5.12}$$

$$g_{4,4} = \int p(X|\theta)\frac{\partial}{\partial\mu_3}\log p(X|\theta)\frac{\partial}{\partial\mu_3}\log p(X|\theta)dx \tag{5.13}$$

For simplicity, we concatenate the components of the mean surface normal $\underline{\mu}$ and write $\underline{\theta} = (\kappa, \underline{\mu}^T)^T$. We perform vector-differentiation with respect to μ to simplify

our calculations.

We commence by computing

$$g_{\kappa,\kappa} = \int f_p(\underline{x}, \kappa, \underline{\mu}) \frac{\partial}{\partial \kappa} \log f_p(\underline{x}, \kappa, \underline{\mu}) \frac{\partial}{\partial \kappa} \log f_p(\underline{x}, \kappa, \underline{\mu}) d\underline{x}. \tag{5.14}$$

Substituting for the vMF distribution, we have

$$g_{\kappa,\kappa} = \int \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa) e^{-\kappa \underline{\mu}^t \underline{x}}}{\kappa^{\frac{p}{2}-1}} \left[\frac{\partial}{\partial \kappa} \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} e^{\kappa \underline{\mu}^t \underline{x}} \right]^2 dx \tag{5.15}$$

$$\begin{aligned} g_{\kappa,\kappa} = \int \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa) e^{-\kappa \underline{\mu}^t \underline{x}}}{\kappa^{\frac{p}{2}-1}} \Bigg[& \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} (\underline{\mu}^t \underline{x}) e^{\kappa \underline{\mu}^t \underline{x}} \\ & + \frac{e^{\kappa \underline{\mu}^t \underline{x}}}{(2\pi)^{\frac{p}{2}}} \left(\frac{I_{\frac{p}{2}-1}(\kappa)(\frac{p}{2}-1)\kappa^{\frac{p}{2}-2} - \frac{1}{2}\kappa^{\frac{p}{2}-1}(I_{\frac{p}{2}-2}(\kappa)) + I_{\frac{p}{2}(\kappa)}}{I_{\frac{p}{2}-1}(\kappa)^2} \right) \Bigg] dx \end{aligned} \tag{5.16}$$

$$\begin{aligned} g_{\kappa,\kappa} = \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)}{\kappa^{\frac{p}{2}-1}} \int e^{\kappa \underline{\mu}^t \underline{x}} & \left(\frac{1}{(2\pi)^{\frac{p}{2}}} \right)^2 \left(\frac{1}{I_{\frac{p}{2}-1}(\kappa)} \right)^2 \left(\kappa^{\frac{p}{2}-2} \right)^2 \\ & \left[\kappa \underline{\mu}^t \underline{x} + \frac{(\frac{p}{2}-1) I_{\frac{p}{2}-1(\kappa)} - \frac{\kappa}{2} \left(I_{\frac{p}{2}-2}(\kappa) + I_{\frac{p}{2}}(\kappa) \right)}{I_{\frac{p}{2}-1}(\kappa)} \right]^2 dx \end{aligned} \tag{5.17}$$

$$g_{\kappa,\kappa} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}} \frac{1}{I_{\frac{p}{2}-1}(\kappa)} \int e^{\kappa \underline{\mu}^t \underline{x}} \left(\frac{1}{\kappa^2}\right) \left[(\kappa \underline{\mu}^t \underline{x})^2 + 2a(\kappa \underline{\mu}^t \underline{x}) + a^2\right] dx \tag{5.18}$$

Finally, we have:

$$g_{\kappa,\kappa} = \frac{1}{\kappa^2} \left(\kappa^2 < cos\theta_\mu >^2 + 2a\kappa < cos\theta_\mu > + a^2\right) \tag{5.19}$$

Where,

$a = \frac{(\frac{p}{2}-1) I_{\frac{p}{2}-1}(\kappa) - \frac{\kappa}{2}\left(I_{\frac{p}{2}-2}(\kappa) + I_{\frac{p}{2}(\kappa)}\right)}{I_{\frac{p}{2}-1}(\kappa)}$

$y = \kappa \underline{\mu}^t \underline{x}$

$e^y = e^{\kappa \underline{\mu}^t \underline{x}}$

$\int y e^y dy = y e^y - \int e^y dy$

$\int y e^y dy = y e^y - e^y$

$\int y e^y dy = e^y (y-1)$

$< cos\theta_\mu >= \int y e^y dy = e^y (y-1)$

With the substitution, we have

$$g_{\kappa,\kappa} = \frac{1}{\kappa^2}\Bigg(\kappa^2(e^{\kappa\underline{\mu}^T\underline{x}}((\kappa\underline{\mu}^T\underline{x})-1)^2) + 2\frac{(\frac{p}{2}-1)I_{\frac{p}{2}-1}(\kappa) - \frac{\kappa}{2}\big(I_{\frac{p}{2}-2}(\kappa)+I_{\frac{p}{2}(\kappa)}\big)}{I_{\frac{p}{2}-1}(\kappa)} \times \kappa(e^{\kappa\underline{\mu}^T\underline{x}}((\kappa\underline{\mu}^T\underline{x})-1))\Bigg) + \left(\frac{(\frac{p}{2}-1)I_{\frac{p}{2}-1}(\kappa) - \frac{\kappa}{2}\big(I_{\frac{p}{2}-2}(\kappa)+I_{\frac{p}{2}(\kappa)}\big)^2}{I_{\frac{p}{2}-1}(\kappa)}\right) \tag{5.20}$$

In the above we can set $p = 3$ since we are dealing with a vMF distribution over a 2D field of surface normals.

In Equation 5.20, we can set $p = 3$ since we are dealing with a vMF distribution over a 2D field of surface normals.

The result of the Equation 5.20 depends on κ, that is, the higher the value of κ, the higher is the value of the Equation 5.20. This is illustrated in Figure 5.1, which shows a sample (from a given image) of values for κ and the resulting value of $g_{\kappa,\kappa}$ – when using the provided value of κ in the calculation. In Figure 5.1 the coordinate x represents κ and the coordinate y represents the result for the equation 5.20 ($g_{\kappa,\kappa}$). The symbol (+) is used to denote the relation between κ and $g_{\kappa,\kappa}$. For example, when $\kappa \cong 12$, $g_{\kappa,\kappa} \cong -0,36$.

The second part is to compute the integral for (5.20), when θ^i and θ^j are equivalent to κ and $\underline{\mu}$, respectively.

$$g_{ij}(\theta) = \int p(X|\theta)\frac{\partial}{\partial\theta^i}\log p(X|\theta)\frac{\partial}{\partial\theta^j}\log p(X|\theta)dx \tag{5.21}$$

$$g_{\kappa,\underline{\mu}} = \int p(X|\theta)\frac{\partial}{\partial\kappa}\log p(X|\theta)\frac{\partial}{\partial\underline{\mu}}\log p(X|\theta)dx \tag{5.22}$$

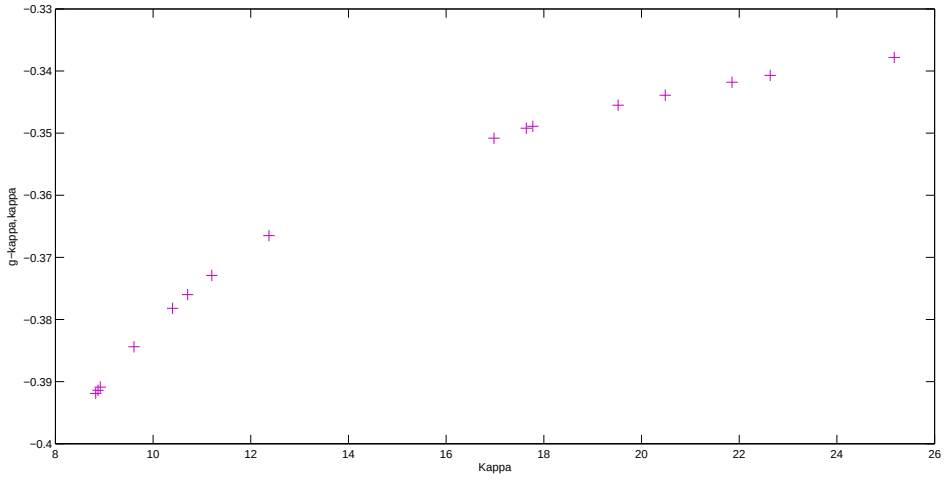

Figure 5.1: Relation between kappa and g-k,k.

$$g_{\kappa,\underline{\mu}} = \int \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)}{\kappa^{\frac{p}{2}-1}} e^{-\kappa \underline{\mu}^t \underline{x}} \left[\frac{\partial}{\partial \kappa} \left(\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} e^{\kappa \underline{\mu}^t \underline{x}} \right) \right] \left[\frac{\partial}{\partial \mu} \left(\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} e^{\kappa \underline{\mu}^t \underline{x}} \right) \right] dx \tag{5.23}$$

$$g_{\kappa,\underline{\mu}} = \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)}{\kappa^{\frac{p}{2}-1}} \int e^{-\kappa \underline{\mu}^t \underline{x}} \left[\frac{e^{\kappa \underline{\mu}^t \underline{x}}}{(2\pi)^{\frac{p}{2}}} \left(\frac{I_{\frac{p}{2}-1}(\kappa)(\frac{p}{2}-1)\kappa^{\frac{p}{2}-2} - \kappa^{\frac{p}{2}-1} \frac{\partial}{\partial \kappa} I_{\frac{p}{2}-1}}{(I_{\frac{p}{2}-1}(\kappa))^2} \right) \right] \left[\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \left(e^{\kappa \underline{\mu}^t \underline{x}} \right) \frac{\partial}{\partial \mu} \kappa \underline{\mu}^t \underline{x} \right] dx \tag{5.24}$$

$$g_{\kappa,\underline{\mu}} = \frac{1}{(2\pi)^{\frac{p}{2}}} \int \frac{I_{\frac{p}{2}-1}(\kappa)(\frac{p}{2}-1)\kappa^{\frac{p}{2}-2} - \kappa^{\frac{p}{2}-1}\left[\frac{1}{2}\left(I_{\frac{p}{2}-2}(\kappa) + I_{\frac{p}{2}}(\kappa)\right)\right]}{(I_{\frac{p}{2}-1}(\kappa))^2}(e^{\kappa\underline{\mu}^t\underline{x}})(\kappa\underline{x})dx \tag{5.25}$$

$$g_{\kappa,\underline{\mu}} = \frac{\kappa}{(2\pi)^{\frac{p}{2}}}\kappa^{\frac{p}{2}-2}\left[\frac{2\left(\frac{p}{2}-1\right)I_{\frac{p}{2}-1}(\kappa) - \kappa\left(I_{\frac{p}{2}-2}(\kappa) - I_{\frac{p}{2}(\kappa)}\right)}{2(I_{\frac{p}{2}-1}(\kappa))^2}\right]\int (e^{\kappa\underline{\mu}^t\underline{x}})\underline{x}dx \tag{5.26}$$

$$g_{\kappa,\underline{\mu}} = \left[\frac{2\left(\frac{p}{2}-1\right)I_{\frac{p}{2}-1}(\kappa) - \kappa\left(I_{\frac{p}{2}-2}(\kappa) - I_{\frac{p}{2}(\kappa)}\right)}{2(I_{\frac{p}{2}-1}(\kappa))^2}\right]\int \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}(e^{\kappa\underline{\mu}^t\underline{x}})\underline{x}dx \tag{5.27}$$

$$g_{\kappa,\underline{\mu}} = \left[\frac{2\left(\frac{p}{2}-1\right)I_{\frac{p}{2}-1}(\kappa) - \kappa\left(I_{\frac{p}{2}-2}(\kappa) - I_{\frac{p}{2}(\kappa)}\right)}{2(I_{\frac{p}{2}-1}(\kappa))^2}\right]\underline{\mu} \tag{5.28}$$

The third part is to compute the integral for (5.20), when θ^i and θ^j are equivalent to $\underline{\mu}$ and $\underline{\mu}$, respectively.

$$g_{ij}(\theta) = \int p(X|\theta)\frac{\partial}{\partial\theta^i}\log p(X|\theta)\frac{\partial}{\partial\theta^j}\log p(X|\theta)dx \tag{5.29}$$

$$g_{\underline{\mu},\underline{\mu}} = \int p(X|\theta)\frac{\partial}{\partial\underline{\mu}}\log p(X|\theta)\frac{\partial}{\partial\underline{\mu}}\log p(X|\theta)dx \tag{5.30}$$

$$g_{\underline{\mu},\underline{\mu}} = \int \frac{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}{\kappa^{\frac{p}{2}-1}}e^{-\kappa\underline{\mu}^t\underline{x}}\left(\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}\right)^2\frac{\partial}{\partial\underline{\mu}}(e^{\kappa\underline{\mu}^t\underline{x}})dx \tag{5.31}$$

$$g_{\underline{\mu},\underline{\mu}} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}\int (e^{-\kappa\underline{\mu}^t\underline{x}})\left[(e^{\kappa\underline{\mu}^t\underline{x}})\kappa\underline{x}\right]^2dx \tag{5.32}$$

$$g_{\underline{\mu},\underline{\mu}} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}\int (\kappa\underline{x})^2e^{\kappa\underline{\mu}^t\underline{x}}dx \tag{5.33}$$

$$g_{\underline{\mu},\underline{\mu}} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}\kappa^2\underline{\mu}^2 \tag{5.34}$$

The result of the Equation 5.37 depends on κ. That is, the higher the value of κ, the lower is the resulting value of the equation 5.20. This is illustrated in Figure 5.2, which shows a sample (from a given image) of values for κ and the resulting value of

$g_{\underline{\mu},\underline{\mu}}$ – when using the provided value of κ in the calculation. Like in Figure 5.1, in the graphic of Figure 5.2 the coordinate x represents κ and the coordinate y represents the result for the Equation, in this case equation 5.37 ($g_{\underline{\mu},\underline{\mu}}$). The symbol (+) is used to denote the relation between κ and $g_{\underline{\mu},\underline{\mu}}$. For instance, when $kappa \cong 11$, $g_{\underline{\mu},\underline{\mu}} \cong 0.07$.

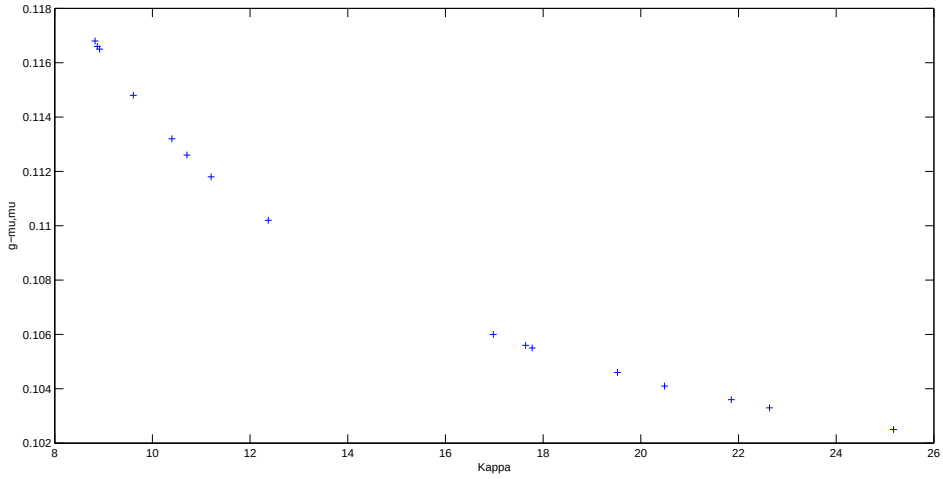

Figure 5.2: Relation between kappa and g-mu,mu.

Finally, we compute

$$g_{\underline{\mu},\underline{\mu}} = \int f_p(\underline{x}, \kappa, \underline{\mu}) \frac{\partial}{\partial \underline{\mu}} \log f_p(\underline{x}, \kappa, \underline{\mu}) \frac{\partial}{\partial \underline{\mu}} \log f_p(\underline{x}, \kappa, \underline{\mu}) d\underline{x}. \tag{5.35}$$

Substituting for the vMF distribution, we have

$$g_{\underline{\mu},\underline{\mu}} = \int \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)}{\kappa^{\frac{p}{2}-1}} e^{-\kappa \underline{\mu}^t \underline{x}} \left(\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \right)^2 \left\{ \frac{\partial}{\partial \underline{\mu}} (e^{\kappa \underline{\mu}^t \underline{x}}) \right\} \left\{ \frac{\partial}{\partial \underline{\mu}} (e^{\kappa \underline{\mu}^t \underline{x}}) \right\}^T d\underline{x}. \tag{5.36}$$

On simplification

$$g_{\underline{\mu},\underline{\mu}} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \kappa^2 \underline{\mu}\mu^T. \tag{5.37}$$

which is a 3x3 matrix.

Substituting for the vMF distribution in (5.2), the elements of the (4×4) Fisher information matrix (M) are:

$$M = \left(\begin{array}{c|c} g_{\kappa,\kappa}^{(1\times1)} & g_{\kappa,\underline{\mu}}^{(1\times3)} \\ \hline g_{\underline{\mu},\kappa}^{(3\times1)} & g_{\underline{\mu},\underline{\mu}}^{(3\times3)} \end{array} \right)$$

We make use of the Fisher-Rao metric to compute the geodesic distance between the two parametric densities. Consider two corresponding 4x4 image regions for which the estimated parameter vectors are:

$$\theta^{a_k} = (\kappa_{a_k}, \underline{\mu}_{a_k})^T.$$

$$\theta^{b_k} = (\kappa_{b_k}, \underline{\mu}_{b_k})^T.$$

$$\hat{\kappa} = \tfrac{1}{2}(\kappa_{a_k} + \kappa_{b_k}).$$

$$\underline{\hat{\mu}} = \tfrac{1}{2}(\mu_{a_k} + \mu_{b_k}).$$

For small changes in parameters the geodesic distance between parameter vectors is:

$$\begin{aligned} ds^2_{a_k,b_k} &= g_{\hat{\kappa},\hat{\kappa}}(\kappa_{a_k} - \kappa_{b_k})^2 + 2(g_{\hat{\kappa},\underline{\hat{\mu}}})^T(\kappa_{a_k} - \kappa_{b_k})(\underline{\mu}_{a_k} - \underline{\mu}_{b_k}) \\ &+ (\underline{\mu}_{a_k} - \underline{\mu}_{b_k})^T g_{\underline{\hat{\mu}},\underline{\hat{\mu}}}(\underline{\mu}_{a_k} - \underline{\mu}_{b_k}). \end{aligned} \tag{5.38}$$

Now, to discover the desired geodesic between two parametric densities, we can use the Fisher-Rao metric (5.29) to calculate the distance between the faces.

$$ds = g_{\kappa,\kappa}(\hat{\kappa_1} - \hat{\kappa_2})^2 + 2g_{\kappa,\underline{\mu}}(\hat{\kappa_1} - \hat{\kappa_2}) \times (\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2) + g_{\underline{\mu},\underline{\mu}}(\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2)^2 \tag{5.39}$$

To compute the total facial dissimilarity, we sum the geodesic distances over all 4x4 non-overlapping image blocks. The total dissimilarity is given by

$$D^2_{a,b} = \sum_k ds^2_{a_k,b_k}. \tag{5.40}$$

5.3 Experimental Results

Our experiments are concerned with assessing shape variation in fields of surface normals due to both facial expression and gender difference. In the case of facial expression we aim to explore the changes in facial shape due to subjects pulling seven different expressions namely, happiness, sadness, surprise, fear, anger, disgust and neutral. We also aim to explore if the techniques described can be used to distinguish the gender of different subjects.

The procedure adopted is as follows. We estimate fields of surface normals by computing the derivatives of the height data, and projecting these onto a fronto-parallel plane. We refer to the fields of surface normals obtained as facial needle-

maps. We align the needle-maps obtained from the different range of images to give the maximum overlap (correlation). Each field of surface normals is tessellated into non-overlapping 4x4 blocks. For each pair of blocks, we estimate the mean surface normal direction and the concentration parameter. For each pair of facial needle-maps be compute the Fisher-Rao metric on a block-by-block basis, and then compute the dissimilarity by summing over the blocks. For the set of n faces under consideration we construct a $n \times n$ dissimilarity matrix. We then apply embedding technique (MDS) to the dissimilarity matrix to obtain embedding co-ordinates for the n faces.

For gender discrimination and face expression, we use MDS, heat kernel and commute time embedding technique to analyze the experiments (details about those techniques 4.2 are described in the Chapter 4). Also, we assess the quality of the resulting low-dimensional data representation by evaluating to what extent the local structure of the data is retained. The evaluation is performed by measuring the Classification error of a 1-Nearest Neighbour (1-NN) classifier that is trained on the low-dimensional data representation. Here an object is simply assigned to the class of its nearest neighbour [11],[103]. In addition, we use the Rand Index to assess the degree of agreement between two partitions of the same set of objects. Based on extensive empirical comparison of several such measures, Milligan and Coooper, 1986 [65] recommendes the Rand Index as the measure of agreement even when comparing partitions having different numbers of clusters [65],[76],[120].

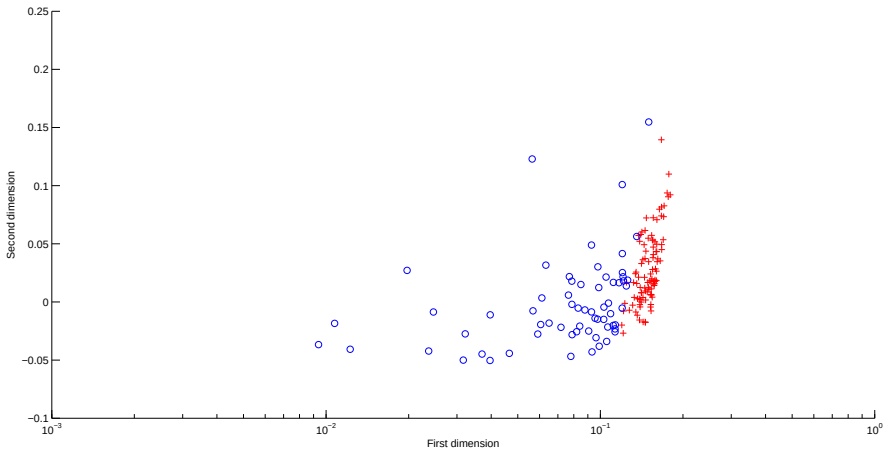

Figure 5.3: Gender difference - Max Planck data set.

5.3.1 Gender Discrimination

We experiment on two sets of data. One is the ground-truth needle-maps calculated from the Max Planck data set. The Max-Planck Face Database [6] [104] comprises 200 (100 females and 100 males) laser scanned (Cyberware TM) human heads without hair. The facial needle-maps are obtained by first orthographically projecting the facial range scans onto a frontal view plane, and then aligning the plane according to the eye centers, and cropping the plane 142x124 pixels to maintain only the inner part of the face. Finally, the surface normal at each pixel position is computed using gradients of the processed range image.

Figures 5.3 and 5.4, show MDS embedding of the pattern of distances into a 2-dimensional space for, respectively, Max Planck data set and EAR data set.

These MDS embedding show the best results achieved using 1-NN classifier. The blue markers are used to denote male subjects, and the red ones female subjects. We can draw the following conclusions from these plots. First, turning our atten-

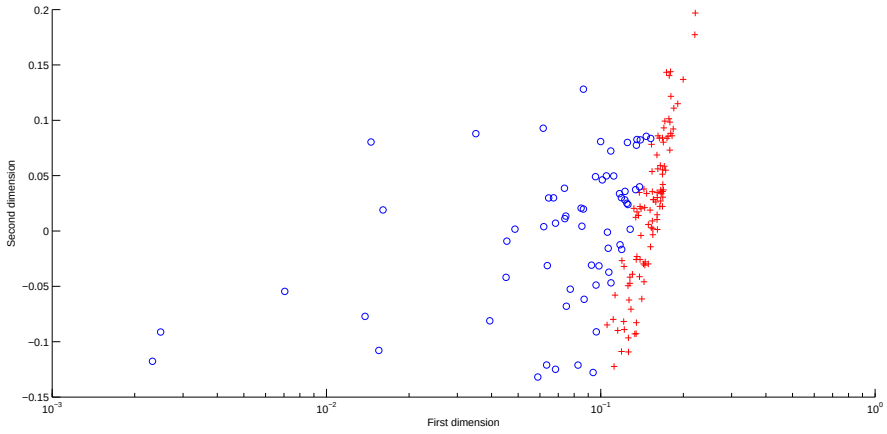

Figure 5.4: Gender difference - EAR data set.

tion to the embedding, using the Fisher-Rao metric the distribution of male and female markers are concentrated differently. In particular the female markers are more densely concentrated. This would suggest that probabilistic separation may be feasible, and the unambiguous male subjects separated from the female ones. Second, it is worth noting that attempting to discriminate male and females faces on the basis of shape alone is a difficult task, and human observers make numerous additional cues such as hair-style.

Table 5.1 shows the results using 1-NN classifier training for MDS, Heat Kernel and Commute Time. The results achieved in MDS gives the best result achieved so far. We have 96% of recognition using Max Plank data set and 97% using EAR data set. Using the other two embedding we got an average of 60% of recognition.

See Table 5.2, we analyse that using Rand Index classifier training we achieved the best results using heat kernel embedding technique, the average for both data set is 80% of recognition.

We observe in Table 5.1 that the performance from the 1-NN classifier gave the best result using MDS embedding technique for both Data sets. Table 5.2 shows that, using Rand Index technique, the best result for both data sets is heat kernel embedding. Also, commute time is not a good classifier for both Classification errors technique.

Table 5.1: Classification error of 1-NN classifier

Embedding	Max Planck data set	EAR data set
MDS	**0.0455**	**0.028**
Heat Kernel	0.4697	0.3333
Commute Time	0.4242	0.4091

Table 5.2: Classification error of Rand Index

Embedding	Max Planck data set	EAR data set
MDS	0.1450	0.3200
Heat Kernel	**0.1200**	**0.1950**
Commute Time	0.4900	0.4900

5.4 Gender Identification Performance

Compared to the results from Section 5.2, it is clear that using Fisher-Rao metric to classify gender difference provide best results. Recognising gender difference is an advance of the research in the field to recognise gender difference. Also, we can compare our results with the work by Wu [112], which developed statistical methods to find gender discriminating features from facial needle-maps. The method constructs a gender sensitive weight maps to quantify the non-uniform distribution, and develop three novel variants of PGA, namely, weighted PGA, supervised

weighted PGA, and supervised PGA. The weight map used in weighted PGA is a straightforward difference between the mean faces of the men and women. The best classification accuracy achieved using supervised weighted PGA is of 92.5%. This accuracy is not only higher than that achieved using standard PGA (87.5%), but also higher than the accuracy of 88.5% achieved using linear discriminant analysis. To improve this weight map construction in supervised weighted PGA by learning the weight map from all the labeled data. Unlike the above two methods, the weight map in supervised PGA describes the pairwise relationship between labeled data. The weight maps is incorporated into the construction of gender discriminating models, and these models are used to extract gender discriminating features. For this method the classification accuracy in the work is of 97%. Also, Lu Xiaoguang et al. [57] proposed a multimodal facial gender and ethnicity identification. Two different modalities of human faces, range and intensity are explored. The range information, containing 3D shape of the face object, is utilized for ethnicity identification; Wen Yi Zhao et al. [124] proposed a method based on shape-from-shading (SFS) which improves the performance of a face recognition system in handling variations due to pose and illumination via image symsthesis. In the Table 5.3, we can observe a comparative with the related works. Ziyi Xu et al. [117] proposed a novel hybrid face coding method by fusing appearance features and geometric features. Volker Blanz et al. [5] presented a method for face recognition across large changes in viewpoint. The method is based on a Morphable method of 3D faces that represents face-specific information extracted from a dataset of 3D scans.

Analyzing the Table 5.3, our research compared to related works achieved the best performance. We achieved $97,20\%$ using EAR data set with classification error

Table 5.3: Gender Identification Performance

Related Works	Gender Identification Performance
Lu Xiaoguang et al. [57]	97%
Wen Yi Zhao et al. [124]	93%
Zing Wu [112]	97%
Ziyi Xu et al. [117]	92, 38%
Volker Blanz et al. [5]	84, 75%

1-NN. Also, we have a success using classification error of Rand Index with 88% of identification using Max Plank data set.

5.4.1 Face expressions

The second experiment explores the ability of the Fisher-Rao metric to distinguish the same face when presented in a different expression. We use the data set from the BU-3DFEDB database [56]. Our experiments consist on a group of 100 faces and 7 expression for each face (neutral, angry, fear, disgust, happy, surprise and sad).

The experiments consist of two parts. Firstly, we start to compute our experiments dividing the BU-3DFEDB database into 11 sets of faces. Then, after doing some testing to validate the technique with the sets of faces, we start working with the whole database (100 faces and 7 expressions). Using this database we select random faces that are compared to the rest of the faces.

Experiments - Part One

This section shows the results using the 11 set of faces from the BU-3DFEDB database. Some visual results of our results are shown in Figure 5.5, Figure 5.6

and Figure 5.7. In our embedding visualizations (MDS) we show the different expressions for the same subject with the same symbol. We have 6 different databases. The first four sets contain male and female subjects. Figure 5.5 shows the first set, which contains 7 faces with 7 expression and the second set with 10 faces with 7 expressions. Figure 5.6 shows the third set containing 15 faces with 7 expressions and the fourth set with 20 faces with 7 expressions. Figure 5.7 shows the final two sets contain, respectively, 7 expressions for 6 males and 6 females.

We observe that the performance from the MDS embedding (see Figure 5.5, Figure 5.6 and Figure 5.7), for a small number of faces in the data-set, achieves good separation of different faces under varying expression. However, as the number of faces increases the overlapping becomes significant. We can see that for data sets with the same gender the results achieve a better performance.

As described previously, we divide the database into sets of faces. Using the sets of faces, we calculate 1-NN and Rand-Index classifier training for each embedding (MDS, heat kernel and commute time). To improve the results achieved so far, we converte the non-metric and non-Euclidean matrix to metric and Euclidean matrix, where we have only non-negatives values [110] [116]. Using the route adopted by Pekalsha [71], by adding a certain constant to all values of the related Gram matrix, and thus compensating for the effect of the negative values (while maintaining the same vector structure). Our aim is to rectify a given set of non-Euclidean dissimilarity data so as to make them more Euclidean. One way to gauge the degree to which a pairwise distance matrix contains non-Euclidean artefacts is to analyse the properties of its centralised Gram matrix. For an $N \times N$ symmetric pairwise dissimilarity matrix D with the pairwise distance as elements, the centered Gram matrix

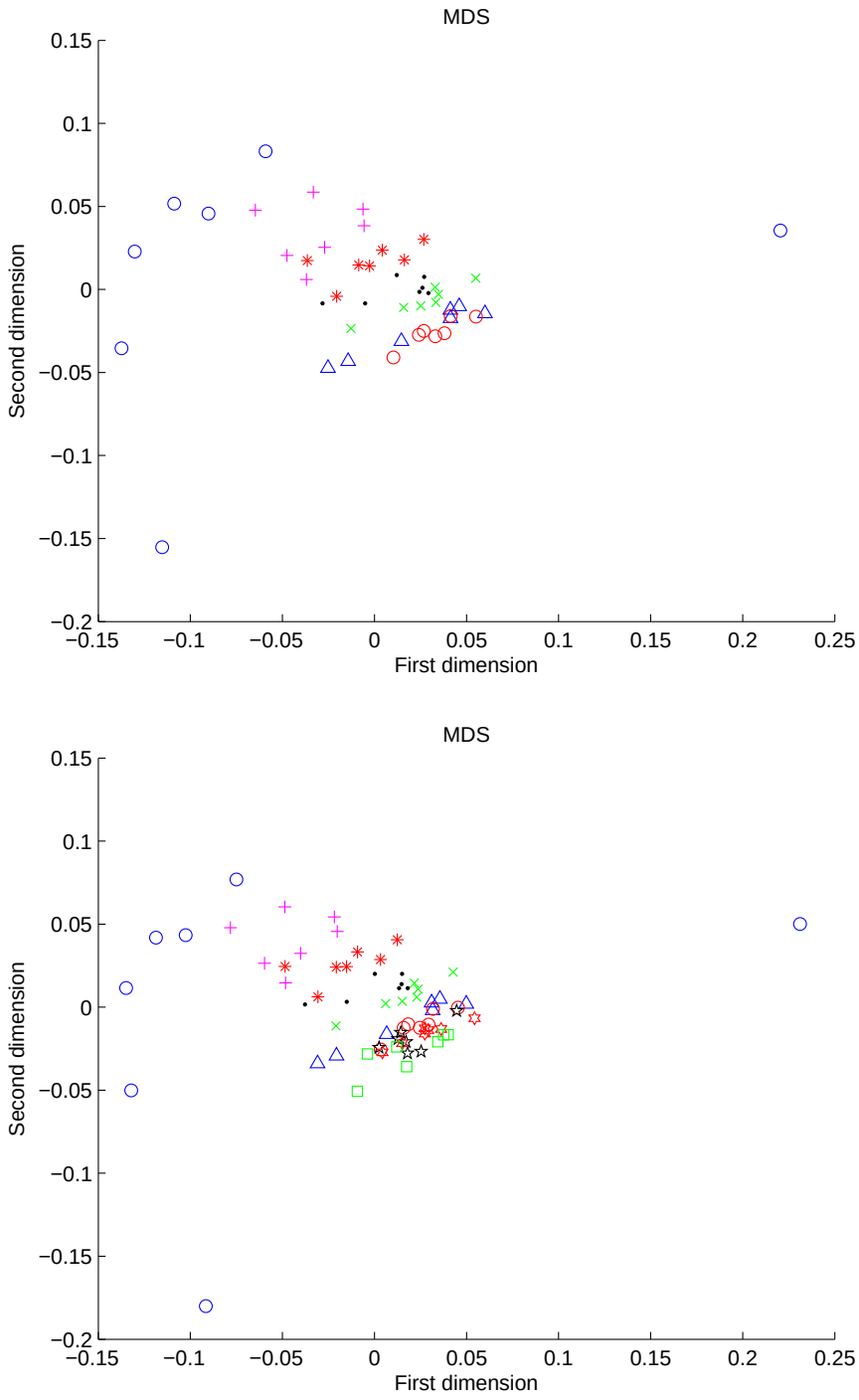

Figure 5.5: In the graphic, at the top embedding consists of 7 different faces each of which appears in 7 different facial expressions. The graphic at the botton consists of 10 different faces each of which appears in 7 different facial expressions..

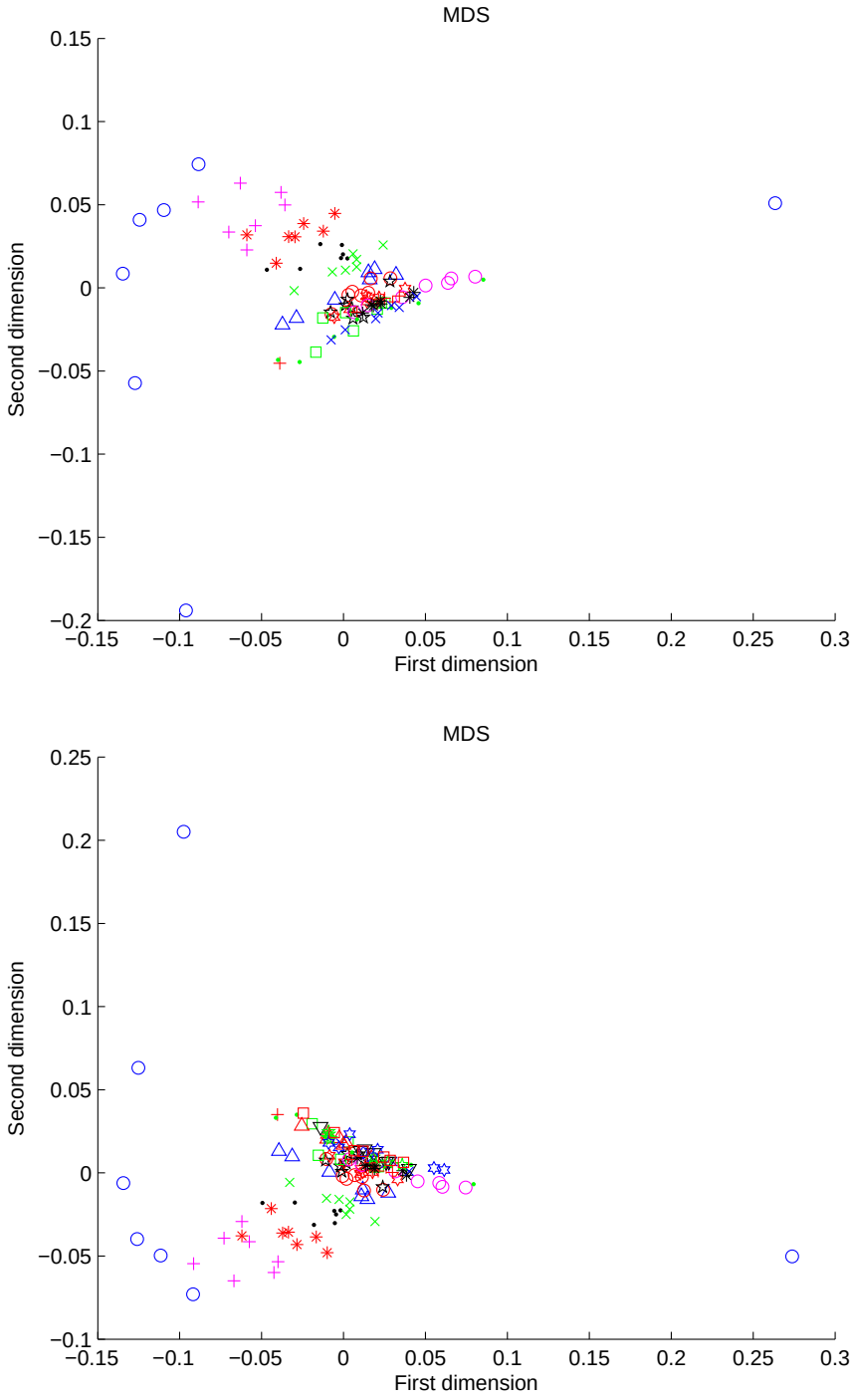

Figure 5.6: In the graphic, at the top embedding consists of 7 different faces each of which appears in 15 different facial expressions. The graphic at the bottom consists of 20 different faces each of which appears in 7 different facial expressions..

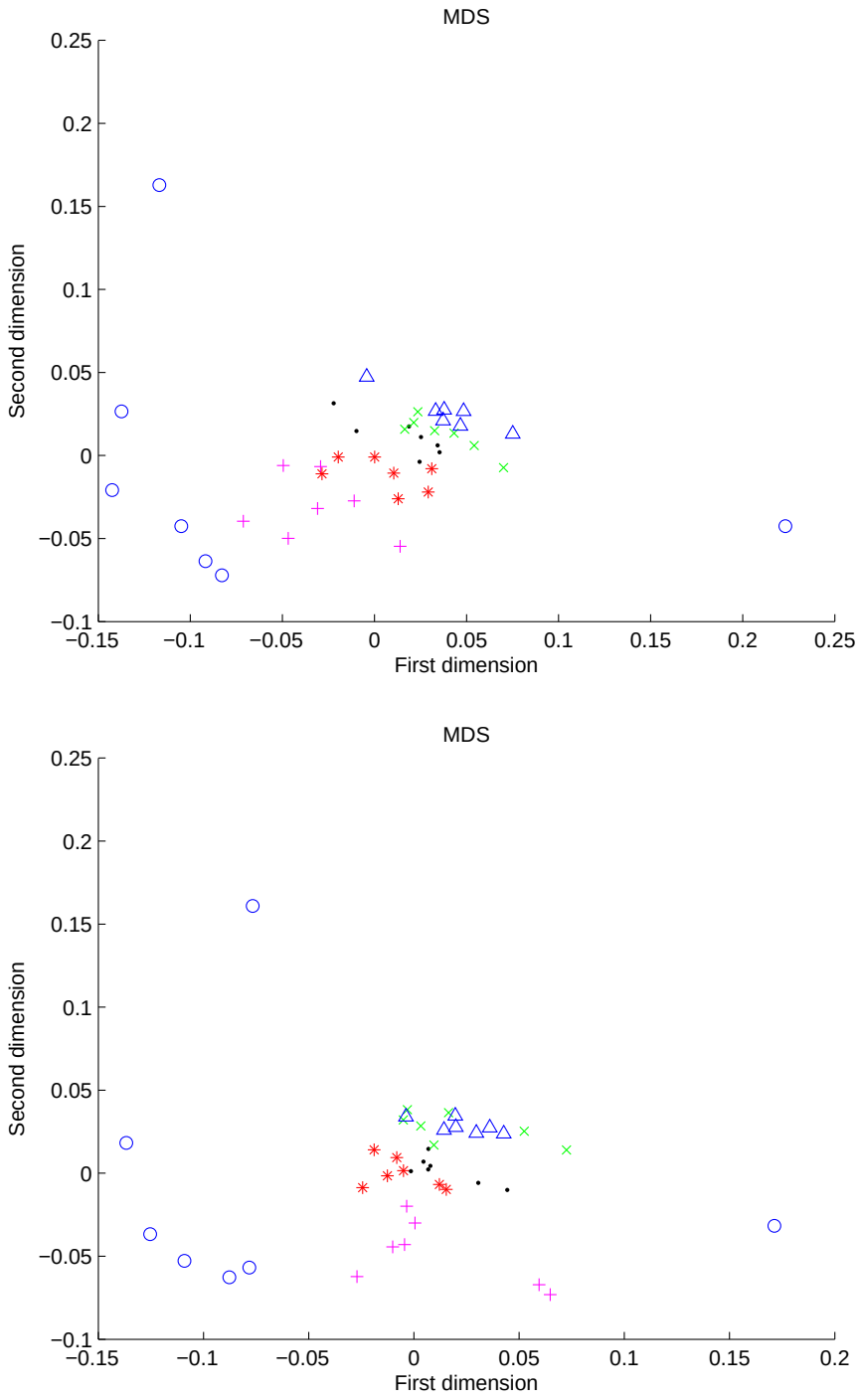

Figure 5.7: In the graphic, at the top embedding consists of 6 different faces each of which appears in 7 different facial expressions, where there are only male faces. The graphic at the bottom consists of 6 different faces each of which appears in 7 different facial expressions, where there are only female faces.

$G = -\frac{1}{2}JD^2J$, where D^2 is element wise squaring of elements in D, $J = I - \frac{1}{N}11^T$ is the centering matrix and 1 is the all-ones vector of length N. The degree to which the distance matrix departs from being Euclidean can be measured by using the relative mass of negative values or negative fraction $\sum_{\lambda<0}|\lambda_i|/\sum_{i=1}^{N}|\lambda_i|$ [71] and [115]. This measure is zero for Euclidean distances and increases as the distance becomes increasingly non-Euclidean. The Kernel embedding is obtained from the centered Gram matrix using the factorisation $G = YY^T$, where Y is the $N \times N$ matrix with the embedded co-ordinates of the data as columns. To determine whether the Gram matrix is positive semi definite [71] and [115], we perform the decomposition $G = \Phi\Lambda\Phi^T$ on the Gram matrix, where $\Lambda = diag(\Lambda_1, ..., \Lambda_N)$ is the diagonal matrix which the ordered values as elements and $\Phi = (\phi_1|...|\phi_N)$ is the vector matrix with the ordered vectors $\phi_1, ..., \phi_N$ as columns. In terms of the values and vectors, the matrix of embedded coordinates is given by $Y = \Phi\sqrt{\Lambda}$ where the values Λ are positive.

In Tables 5.4, 5.5 and 5.6, respectively, MDS, Heat Kernel and Commute time, we can see the new values using 1-NN classifier training, the non-negative values achieved a better performance compared to negative values.

In Tables 5.7, 5.8 and 5.9, respectively, MDS, Heat Kernel and Commute Time embedding techniques, using Rand Index classifier training are shown.

The difference between the results of the experiments using the 1-NN and the Rand Index classifier trained are: for a database with a mixture of male and female the best results achieved was using MDS for both classifier training. Also, for MDS, heat kernel and commute time, when the set of faces starts to increase, the results start to increase for both cases, using negative eigenvalues and non-negative

Table 5.4: Classification errors of 1-NN classifier trained using MDS embedding.

Face Expressions/DataBase	neg.eigenvalues	non-neg.eigenvalues
7 Faces 7 Expressions	0.1633	0.0612
10 Faces 7 Expressions	0.2857	0.0714
15 Faces 7 Expressions	0.3429	0.0952
20 Faces 7 Expressions	0.4429	0.1571
30 Faces 7 Expressions	0.5286	0.1905
40 Faces 7 Expressions	0.6710	0.1964
50 Faces 7 Expressions	0.6710	0.2400
68 Faces 7 Expressions	0.7563	0.3130
71 Faces 7 Expressions	0.7606	0.3038
6 Faces 7 Expressions (Male Expressions)	0.1905	0.1190
6 Faces 7 Expressions (Female Expressions)	0.2619	0.0476

Table 5.5: Classification errors of 1-NN classifier trained using Heat Kernel embedding.

Face Expressions/DataBase	neg.eigenvalues	non-neg.eigenvalues
7 Faces 7 Expressions	0.8776	0.7143
10 Faces 7 Expressions	0.9000	0.8571
15 Faces 7 Expressions	0.9143	0.9238
20 Faces 7 Expressions	0.8929	0.9286
30 Faces 7 Expressions	0.5190	0.9429
40 Faces 7 Expressions	0.9464	0.9229
50 Faces 7 Expressions	0.9571	0.9657
68 Faces 7 Expressions	0.9790	0.9706
71 Faces 7 Expressions	0.9618	0.9678
6 Faces 7 Expressions (Male Expressions)	0.8095	0.7619
6 Faces 7 Expressions (Female Expressions)	0.7857	0.7857

Table 5.6: Classification errors of 1-NN classifier trained using Commute Time embedding.

Face Expressions/DataBase	neg.eigenvalues	non-neg.eigenvalues
7 Faces 7 Expressions	0.7755	0.8333
10 Faces 7 Expressions	0.8429	0.9000
15 Faces 7 Expressions	0.9048	0.9333
20 Faces 7 Expressions	0.9214	0.9500
30 Faces 7 Expressions	0.9667	0.9697
40 Faces 7 Expressions	0.9750	0.9750
50 Faces 7 Expressions	0.9800	0.9800
68 Faces 7 Expressions	0.9667	0.9853
71 Faces 7 Expressions	0.9859	0.9859
6 Faces 7 Expressions (Male Expressions)	0.7619	0.8333
6 Faces 7 Expressions (Female Expressions)	0.7381	0.8333

Table 5.7: Classification errors of Rand Index classifier trained using MDS embedding.

Face Expressions/DataBase	neg.eigenvalues	non-neg.eigenvalues
7 Faces 7 Expressions	0.2041	0.2653
10 Faces 7 Expressions	0.2571	0.2000
15 Faces 7 Expressions	0.3163	0.1619
20 Faces 7 Expressions	0.4662	0.1714
30 Faces 7 Expressions	0.6429	0.1762
40 Faces 7 Expressions	0.7000	0.2179
50 Faces 7 Expressions	0.7543	0.2000
68 Faces 7 Expressions	0.8088	0.2542
71 Faces 7 Expressions	0.8169	0.2515
100 Faces 7 Expressions	0.8249	0.3702
6 Faces 7 Expressions (Male Expressions)	0.1190	0.2619
6 Faces 7 Expressions (Female Expressions)	0.1667	0.2857

Table 5.8: Classification errors of Rand Index classifier trained using Heat Kernel embedding.

Face Expressions/DataBase	neg.eigenvalues	non-neg.eigenvalues
7 Faces 7 Expressions	0.2857	0.2653
10 Faces 7 Expressions	0.3857	0.3143
15 Faces 7 Expressions	0.4694	0.3238
20 Faces 7 Expressions	0.6090	0.3643
30 Faces 7 Expressions	0.7748	0.3333
40 Faces 7 Expressions	0.7786	0.2929
50 Faces 7 Expressions	0.8429	0.3143
68 Faces 7 Expressions	0.8697	0.3887
71 Faces 7 Expressions	0.8753	0.4125
100 Faces 7 Expressions	0.8793	0.8793
6 Faces 7 Expressions (Male Expressions)	0.2381	0.4048
6 Faces 7 Expressions (Female Expressions)	0.2857	0.3810

Table 5.9: Classification errors of Rand Index classifier trained using Commute Time embedding.

Face Expressions/DataBase	neg.eigenvalues	non-neg.eigenvalues
7 Faces 7 Expressions	0.4082	0.3061
10 Faces 7 Expressions	0.5571	0.3714
15 Faces 7 Expressions	0.6122	0.3714
20 Faces 7 Expressions	0.6692	0.4000
30 Faces 7 Expressions	0.7714	0.3143
40 Faces 7 Expressions	0.8214	0.3214
50 Faces 7 Expressions	0.8657	0.3286
68 Faces 7 Expressions	0.8866	0.3992
71 Faces 7 Expressions	0.8913	0.3984
100 Faces 7 Expressions	0.8793	0.8793
6 Faces 7 Expressions (Male Expressions)	0.3810	0.4048
6 Faces 7 Expressions (Female Expressions)	0.3810	0.4762

eigenvalues. For databases where we have only one subject, female or male, generally, is 1-NN classifier training using MDS embedding (see Table 5.4). Otherwise, using the negative eigenvalues we have another analysis where for database where we have only one subject, female or male, the best classifier trained is Rand Index using MDS (see Table 5.7) and for databases with mixtures of male and female we have the same classifier trained, 1-NN, where we achieved the best results (see Table 5.4).

Experiments - Part Two

Working with 100 faces (each face has 7 expressions, a total of 700 faces in the database), we calculate the 1-NN error using a test data set that is randomly selected from the 700 faces for 10.000 iterations. We set the range of the faces in the test set from 1 to 7.

Tables 5.10 to 5.30 show the results for each classifier (MDS, Heeat Kernel and Commute Time), respectively, with the number of the faces in the test set. Like we did in the previous section (Experiments - Part one), we compare the results using the non-negative values. Next, we show the results for each randomly selected set of faces from the database.

Random Testing with 1 face

Tables 5.10 , 5.11 and 5.12 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques. Experiments using only 1 randomly selected face and 699 training faces for 10.000 iterations.

The best result achieved was 0.0 for all the embedding techniques using neg-

ative eigenvalues and non-negative eigenvalues. It means that in some iterations the testing face was recognised into the training data set that contained the other expressions for the same face. The worst case was 1.0 for all embedding techniques, meaning that the face expression was not recognise.

We can see that all the embedding techniques achieved the same best result, but the best average was achieved using MDS compared to the other two embedding technique. Also, using non-negative eigenvalues gave a better average of error 0.5910 of error (0.4090 of recognition). Analyzing the three embedding, we can see that the disgust and angry expressions gave the best recognition result. On the other hand, the worst case shows that fear, in this test, is the face expression that is most difficult to recognise in the database.

Random Testing with 2 faces

Tables 5.13, 5.14, 5.15 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques for experiments using 2 randomly selected faces and 698 training faces for 10.000 iterations.

The best achieved result was 0.0 for MDS and Heat Kernel embedding techniques using negative eigenvalues. For Commute Time it achieved 0.5 of error for negative and non-negative eigenvalues. Also, the worst recognition case was 1.0 for all embedding techniques.

The MDS embedding technique achieved a best result for negative and non-negative eigenvalues. Analyzing the three embedding, we can see when using MDS embedding, the best and worst face recognition is for the same expression - Fear.

Table 5.10: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 1 Face	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.0	0.0
Best Faces	Disgust	Angry
Worst Result	1.0	1.0
Worst Faces	Fear	Fear
Average	0.8110	0.5910

Table 5.11: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 1 Face	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.0	0.0
Best Faces	Angry	Angry
Worst Result	1.0	1.0
Worst Faces	Fear	Fear
Average	0.9830	0.9840

Table 5.12: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 1 Face	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.0	0.0
Best Faces	Disgust	Angry
Worst Result	1.0	1.0
Worst Faces	Fear	Fear
Average	0.9890	0.9930

For Heat Kernel and Commute Time show, that Neutral expression gave the best and worst recognition results.

Random Testing using 3 Faces

Tables 5.16, 5.17, 5.18 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques for experiments using 3 randomly faces and 697 training faces for 10.000 iterations.

The best result achieved was 0.0 for MDS embedding technique using negative eigenvalues and non-negatives eigenvalues. For Heat Kernel achieved 0.3333 using non-negative eigenvalues and Commute Time achieved 0.3333 of error using negative eigenvalues. It means that for some iterations the face testing was recognise into the training data set that contain the other expressions from the same face. Also, we have the worst case 1.0 for all embedding techniques.

The MDS embedding technique achieved a best result for negative and non-negative eigenvalues. Analyzing the three embedding, we can see that in all the embedding techniques, for the best and worst faces to recognise it is not significant to analyse what expression helps to find the good results. The best average result is using MDS with 0.5746 of error.

Random Testing with 4 Faces

Tables 5.19, 5.20, 5.21 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques for experiments using 4 randomly faces and 696 training faces for 10.000 iterations.

Table 5.13: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 2 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.0	0.0
Best Faces	Surprise and Fear	Happy and Fear
Worst Result	1.0	1.0
Worst Faces	Neutral and Fear	Neutral and Fear
Average	0.8325	0.6010

Table 5.14: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 2 Faces	neg.eigenvalues	non-neg.eigenvalues
hline Best Result	0.0	0.50
Best Faces	Neutral	Surprise and Happy
Worst Result	1.0	1.0
Worst Faces	Neutral and Fear	Neutral and Fear
Average	0.9770	0.9890

Table 5.15: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 2 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.50	0.50
Best Faces	Neutral	Neutral and Happy
Worst Result	1.0	1.0
Worst Faces	Neutral and Fear	Neutral and Happy
Average	0.9885	0.9890

Table 5.16: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 3 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.0	0.0
Best Faces	Surprise and Disgust	Happy, Disgust and Neutral
Worst Result	1.0	1.0
Worst Faces	Disgust, Angry and Neutral	Surprise and Disgust
Average	0.8166	0.5746

Table 5.17: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 3 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.6666	0.3333
Best Faces	Surprise, Angry and Fear	Angry and Happy
Worst Result	1.0	1.0
Worst Faces	Surprise, Disgust and Fear	Fear, Surprise and Disgust
Average	0.9793	0.9756

Table 5.18: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 3 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.3333	0.6666
Best Faces	Surprise, Happy and Angry	Neutral, Happy and Disgust
Worst Result	1.0	1.0
Worst Faces	Surprise, Fear and Fear	Fear, Surprise and Disgust
Average	0.9940	0.99

When the size of the face testing set starts to increase, the best results start to decrease. Still, we can find the best result using MDS embedding technique for non-negative eigenvalues. The best average is also achieved using MDS embedding with 0.5567.

Random Testing with 5 faces

Tables 5.22, 5.23, 5.24 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques for experiments using 5 randomly faces and 695 training faces for 10.000 iterations.

The best result is using MDS embedding for non-negative eingvalues. Also, the best average is using MDS with 0.5722. Another analysis is when the set of faces testing starts to increase, the average in kept constant starting with 2 faces testing, using all the embedding techniques. What changes frequently is the best and worst results. As described before, with faces expression does not matter to help finding the best result.

Random Testing with 6 faces

Tables 5.25, 5.26, 5.27 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques for experiments using 6 randomly faces and 694 training faces for 10.000 iterations.

MDS is still doing the best embedding for achieving the best result. In this case, we achieved the best results using non-negative eigenvalues. The average is still constant, using MDS with 0.5765 of error.

Table 5.19: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 4 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.2500	0.0
Best Faces	Happy, Fear and Surprise	Surprise and Angry
Worst Result	1.0	1.0
Worst Faces	Surprise and Angry	Happy
Average	0.8185	0.5767

Table 5.20: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 4 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.5000	0.5000
Best Faces	Surprise, Fear, Neutral and Disgust	Disgust, Fear and Angry
Worst Result	1.0	1.0
Worst Faces	Surprise, Angry and Fear	Surprise, Fear and Angry
Average	0.9760	0.9825

Table 5.21: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 4 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result	0.5000	0.7500
Best Faces	Fear, Surprise, Neutral and Disgust	Surprise, Disgust and Neutral
Worst Result	1.0	1.0
Worst Faces	Surprise and Fear	Surprise, Angry and Fear
Average	0.9760	0.9902

Table 5.22: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 5 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.2000 Fear, Disgust and Angry	0.0 Happy, Disgust and Fear
Worst Result Worst Faces	1.0 Fear, Disgust and Angry	1.0 Disgust, Happy and Angry
Average	0.8216	0.5722

Table 5.23: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 5 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.4000 Surprise, Fear and Neutral	0.6000 Happy, Fear, Disgust and Surprise
Worst Result Worst Faces	1.0 Neutral, Surprise, Disgust and Angry	1.0 Neutral, Surprise, Disgust and Angry
Average	0.9804	0.9900

Table 5.24: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 5 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.6000 Fear, Surprise, Neutral and Angry	0.6000 Angry, Surprise and Fear
Worst Result Worst Faces	1.0 Angry	1.0 Neutral, Surprise Disgust and Angry
Average	0.9902	0.9844

Table 5.25: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 6 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.3333 Happy, Angry and Surprise	0.0 Happy, Surprise, Neutral
Worst Result Worst Faces	1.0 Fear, Neutral, Happy and Angry	1.0 Surprise, Neutral
Average	0.8196	0.5765

Table 5.26: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 6 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.6666 Fear, Happy, Angry and Surprise	0.6666 Happy, Angry, Fear and Disgust
Worst Result Worst Faces	1.0 Fear, Surprise, Neutral Neutral and Fear	1.0 Fear, Surprise and Neutral
Average	0.9798	0.9790

Table 5.27: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 6 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.6667 Happy, Surprise, Neutral and Angry	0.6666 Surprise, Angry, and Disgust
Worst Result Worst Faces	1.0 Fear, Surprise and Neutral	1.0 Fear, Surprise and Neutral
Average	0.9915	0.9903

Random Testing with 7 Faces

Tables 5.28, 5.29, 5.30 show, respectively, MDS, Heat Kernel and Commute Time embedding techniques for experiments using 7 randomly faces and 693 training faces for 10.000 iterations.

We see similar results compared to the other set of testing faces. The best result is 0.0 using MDS embedding and the average is 0.5573, also using MDS embedding. Heat Kernel and Commute Time show a bad average using both negative and non-negative eigenvalues.

5.5 Summary

This chapter presented how Fisher-Rao metric can be used to measure different facets of facial shape estimated from fields of surface normals using the von-Mises Fisher (vMF) distribution. The Fisher information of the family defines on the manifold a Riemannian metric in known as the Fisher-Rao metric. The Fisher-Rao metric has an approximation which is accurate if the variance of the measured error is small. It is shown that the manifold of parameter values has a finite volume under the approximating metric.

Our experiments concerned with assessing shape variation in fields of surface normals due to both facial expression and gender difference. In the case of facial expression, we explored the changes in facial shape due to subjects pulling seven different expressions namely, happiness, sadness, surprise, fear, anger, disgust and neutral. Also we explored how the fisher-rao metric technique could be used to distinguish the gender of different subjects.

Table 5.28: Classification errors of 1-NN using MDS embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 7 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.2857 Happy, Disgust, Neutral, Fear and Surprise	0.0 Fear, Surprise, Disgust and Neutral
Worst Result Worst Faces	1.0 Fear, Surprise, Neutral, Happy and Angry	1.0 Angry, Happy, Fear, Surprise and Neutral
Average	0.8205	0.5573

Table 5.29: Classification errors of 1-NN using Heat Kernel embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 7 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.7142 Neutral, Happy and Surprise and Surprise	0.7142 Fear, Disgust, Neutral and Happy
Worst Result Worst Faces	1.0 Angry, Happy Fear and Neutral	1.0 Fear, Surprise, Disgust Neutral and Happy
Average	0.9778	0.9801

Table 5.30: Classification errors of 1-NN using Commute Time embedding for 10.000 iterations using 100 faces with 7 expressions.

Testing - 6 Faces	neg.eigenvalues	non-neg.eigenvalues
Best Result Best Faces	0.7142 Fear, Happy and Surprise	0.7142 Neutral, Fear, Surprise, Angry and Happy
Worst Result Worst Faces	1.0 Fear, Surprise, Disgust Neutral and Happy	1.0 Fear, Surprise, Disgust, Neutral and Happy
Average	0.9922	0.9898

Analyzing the experiments for face expressions and gender difference, we achieved the best performance for gender difference using Fisher-Rao metric. Using MDS with 1-NN classifier training the average of 0.0303 of error (EAR data set) and 0.0455 of error (Max Planck data set), were achieved. Also, using Rand Index we achieved the best results using Heat Kernel embedding, the average of 0.12 error (Max Planck data set) and 0.1950 error (EAR data set).

On the other hand, for face expression we have a mixture of results. Using the first part of the experiments (divided the BU-3DFEDB into sets of face), the difference between the results of the experiments using the 1-NN and the Rand Index classifier trained are: for database with mixture of male and female the best results achieved was using MDS for both classifier training and the best average was using non-negatives eigenvalues (7 faces and 7 expressions with 0.0612 of error, 10 faces and 7 expressions with 0.0714 of error, 15 faces and 7 expressions with 0.0952 of error, 20 faces and 7 expressions with 0.1571 of error, 30 faces and 7 expressions with 0.1905 of error, 40 faces and 7 expressions with 0.1964 of error, 50 faces and 7 expressions with 0.24, 68 faces and 7 expressions with 0.3130, 71 faces and 7 expressions with 0.3038 and 6 faces and 7 expressions for female and male, respectively, 0.0476 and 0.1190).

Using the second part of the experiments (working with 100 faces and 7 expressions, a total of 700 faces in the database), we achieved different averages using a test data set that is randomly selected from the 700 faces for 10.000 iterations. One conclusion is the same for all the random cases, we achieved the best results, average 0 of error for all MDS embedding and the worst case was of 100% of error. Then, analysing each case, we have for 1 random face: the best average of error

recognition is 0.5910 using non-negative eigenvalues for MDS embedding. For 2 random faces: the best average of error is 0.6010 with non-negative eigenvalues for MDS embedding. For 3 random faces: the best average of error recognition is 0.5746 using non-negative eigenvalues for MDS. For 4 random faces: the best average of error is 0.5767 using non-negative eigenvalues for MDS embedding. For 5 random faces: the best average of error recognition is 0.5722 using non-negative eigenvalues for MDS. For 6 random faces: the best average of error is 0.5765 using non-negative eigenvalues for MDS embedding. And for 7 random faces: the best average of error recognition is 5573 using non-negative eigenvalues for MDS. It is clear that, from the third random faces the average stabilize with 0.57 of error using non-negative eigenvalues and MDS embedding.

CHAPTER 6

Conclusions

In this chapter, we first summarize the main contributions of this thesis, and then analyze the weaknesses of the developed methods. Following the analysis, we discuss some possible solutions and give suggestions for future face expression recognition and gender difference research. Section 6.1 shows a summary of the contributions from the research. Section 6.2, summarize and discuss some problems that are not solved in this thesis. Finally, Section 6.3 shows future research directions motivated by the work.

6.1 Summary of Contributions

The overall goal of this thesis was to use statistical shape analysis to construct shape-spaces that span facial expressions and gender difference by facial needle-maps, and use the resulting shape-model to perform face recognition under varying expression and gender. Facial needle-maps describe the local orientation of facial surfaces, which on one hand reveal the facial shape information, and on the other hand can be recovered from 2D images using shape-from-shading.

In Chapter 4, we presented our first attempt to use facial needle-maps recovered using Shape-From-Shading (SFS) for face expression and gender difference. The results demonstrate the feasibility of face expression. We start using a measure distance to discover the similarity between faces expression and gender difference, which is measured using a number of alternatives including geodesic, Euclidean, Mahalanobis and cosine distance between points on the manifold. In a recognition experiment we compared the perfomance distance between Euclidean, cosine and geodesic distance associated with the shape manifold. We explored if the distances used to distinguish gender and recognise the same face under different expressions.

Chapter 5 explored whether the Fisher-Rao metric can be used to characterise the shape changes due to differences in expression and gender difference. We used a 2.5D representation based on facial surface normals (or facial needle-maps) for gender classification and face expression. The needle-map is a shape representation which can be acquired from 2D intensity images using shape-from-shading (SFS). Using the von-Mises Fisher distribution, we computed the elements of the Fisher information matrix, and used this to compute geodesic distance between fields of surface normals

to construct a shape-space. We embed the fields of facial surface normals into a lower dimensional pattern space using a number of alternative methods including multidimensional scaling, heat kernel embedding and commute time embedding.

6.2 Weaknesses

There are a some of question that are not yet fully solved in this thesis. We summarize and discuss these weaknesses in this section. Some of these weaknesses could be addressed in future work.

The results in Chapter 4 shows how different distance measures can be used to measure different facets of facial shape including gender and expressions. We computed geodesic, Euclidean, Mahalanobis and cosine distances for the long-vector representations of the needle-maps of faces. Then, we visualise the distribution of distances using multidimensional scaling [54] to embed the faces in a two-dimensional pattern space (details of MDS were given in Section 4.2).

The similarity between faces is measured by the geodesic distance on the shape manifold using information provided by surface normals. For the computation of the geodesic distance on shape-space we implemented the method proposed by Small [87] and compared the results to other three measures Euclidean, Mahalanobis and cosine.

In the first contribution of this thesis, we could see that using the method proposed by Small [87] we did not achieve a satisfactory result for face expression and gender difference in our experiments. We can see that in the face expression experiments, only for small numbers of faces data set, we achieved good separation

of different faces under varying expression. In fact, the best separation is achieved with geodesic distance and Mahalanobis distances. However, as the number of faces increases the overlapping becomes significant. It is clear that in the case of the Euclidean and cosine distances, while the distance distributions are overlapping, the mean interclass distance is less than the mean between class distance. In fact, the best separation is given by the cosine and Mahalanobis distances. For the geodesic distance this is not the case, and the distributions are almost indistinguishable.

Using Fisher-Rao metric (Chapter 5) to recognise face expression using sets of faces from the BU-3DFEDB database, we did not achieved a satisfactory results using Heat Kernel and Commute Time. Both classifier training gave an average of 97% error. It is poor results compared with the results of MDS embedding.

6.3 Future Research Directions

Having addressed the weaknesses of this thesis, in this section we state our future work to improve the results presented, and give some sugestions for future research in face expression recognition and gender difference.

There are clearly a number of ways in which this work may be extended. First, we have concentrated on frontal view facial surfaces and we can only recover facial shapes from images with the same viewpoint. We calculate the four distances (Chapter 4) and Fisher-Rao metric (Chapter 5) using 2.5D frontal images from the 3DFEDB database [56], Max-Planck face database [6] and Notre Dame biometric database [36]. Even though the ability to deal with varying pose is almost always claimed to be a benefit of 3D face capture, this problem could be solved by incor-

porating our methods into a set of view-based models similar to those proposed by Pentland et al. [74] and Reisfeld et al. [77] which proposed that a set of separate submanifolds can be obtained by applying PGA to facial needle-maps of each viewpoint.

Second, there is a need to investigate the use of more sophisticated methods such as how the Spherical Triangle can be used to compute the similarity of fields of surface normals. The aim is to show how to compute the areas of geodesic and Euclidean triangles, and use the area ratio as an invariant for the purposes of matching to construct individual shape-spaces for each class of object from interpolation of each triangle as a mean of computing face expressions and gender similarity. To construct the spherical triangle, we have:

- T is the triangulation of a smooth surface S in $\Re^3$;
- Ag is the area of geodesic triangle on S with angles $\alpha_{i}{}_{i=1}^{3}$ and geodesic edge lengths $(d_{gi})_{i=1}^{3}$ and A_e is the area of corresponding Euclidean triangle with lengths $(d_{ei})_{i=1}^{3}$ and angle $\varphi_{i=1}^{3}$;
- assuming that each geodesic is a great arc on a sphere with radius R_i, $i = 1, 2, 3$ corresponding to a central angle 2θ, and that the geodesic triangle is a triangle on a sphere with $R = \frac{1}{3}\sum_{i=1} R$ with the Euclidean distance between the pair of nodes is $d_e = \frac{1}{3}\sum_{i=1}^{3} d_{ei}$;
- considering a small area element on the sphere given in spherical coordinates by $dA = R^2 \sin\theta d\theta d\varphi$, the integration of dA bounded by 2θ gives us the area of the geodesic triangle:

$A = \frac{1}{2R}(dg)(de)^2$

where

$dg(u,v) = 2r_s(u,v)\theta(u,v)$

$de(u,v) = 2r_s(u,v)\sin\theta(u,v)$

- now we can describe the data in terms of triangular faces and assign to each face a curvature ratio K and the Gaussian curvature associated with each vertex will be $G_K = 3dg - \frac{72(dg-de)}{(dg)^3}$

We have started investigating this method and to complete this idea we need to research how to construct the interpolation between the faces using the spherical triangle.

For face expression we did the experiments using a mixture of male and female subjects and only one subject (female face expressions or male expressions). For a subject, we started to do experiments using a small group of faces and we achieved a satisfactory face expression recognition. Also, the set of faces did not prove that the methods are robust for this type of analysis (male face expressions and female face expressions).

Third, we can research about the data set that exists in the literature that could offer face expressions with a number of expressive face expressions for each subject (male and female). Then, we can use the methods provided in this thesis (Chapters 4 and 5) to classify and compare the results.

Bibliography

[1] X. Bai, R. Wilson, and E. R. Hancock. Manifold embedding of graphs using heat kernel. *Mathematics of Surface XI*, 3604(417):34–39, 2005.

[2] S. Baluja and H. and Rowley. Boosting sex identification performance. *International Journal of Computer Vision*, 7(11), January 2007.

[3] P.N. Belhumeur, J.P. Hespanha, and D.J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[4] A. Blake and M. Isard. *Active Contours*. 1998.

[5] V. Blanz, P. Grother, P.J. Phillips, and T. Vetter. Face recognition based on frontal views generated from non-frontal images. *CVPR*, 2:454–461, 2005.

[6] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. *In Proc. SIGGRAPH'99*, pages 187–194, 1999.

[7] F. L. Bookstein. A statistical method for biological shape comparisons. *Journal of theoretical biology*, 107:475–520, 1984.

[8] F. L. Bookstein. Size and spaces for landmark data in two dimensions. *Statistical Science*, 1(2):181–222, 1986.

[9] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[10] F.L. Bookstein. The measurement of biological shape and shape change. *Lecture Notes in Biomathematics.*, 24, 1978. Berlin: Springer.

[11] D. Bremmer, E. Demaine, J. Erickson, J. Lacono, S. Langerman, P. Morin, and G. Toussaint. Output-sensitive algorithms for computing nearest-neighbor decision boundaries. *Discrete and Computational Geometry*, 33 (4):593–604, 2005.

[12] A.M. Bronstein, M.M. Bronstein, and R. Kimmel. Expression-invariant face recognition via spherical embedding. *Conference on Image Processing (ICIP)*, pages 756–759, 2005.

[13] S. Buchala, N. Davey, R.J Frank, and T.M Gale. Dimensionality reduction of face images for gender classification. *Technical Report 408. Department of Computer Science, The University of Hertfordshire, UK*, 2004.

[14] S. Buchala, N. Davey, T.M Gale, and R.J Frank. Principal component analysis of gender, ethnicity, age, and identity of face image. In *IEEE ICMI*, 2005.

[15] A.M. Burton, V. Bruce, and N. Dench. What's the difference between men and women? evidence from facial measurement. *Perception*, 22:153–176, 1993.

[16] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1977.

[17] M. Castelan and E.R. Hancock. Acquiring height data from a single image of a face using local shape indicators. *Computer Vision and Image Understanding*, 1(103):64–79, 2006.

[18] K. Chang, K.W. Boyer, and P.J. Flynn. Face recognition using 2d and 3d facial data. *ACM workshop on multimodal user authentication*, pages 25–32, December 2003.

[19] G. Charpiat, O. Faugeras, and Keriven. R. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Found. Comput. Math.*, 5(1):1 – 58, 2005.

[20] L. S Chen, H. Tao, T.S. Huang, T. Miyasato, and R. Nakatsu. Emotion recognition from audiovisual information. *In Proc. IEEE Workshop on Multimedia Signal Processing*, Los Angeles, CA, USA.:83–88, December 1998.

[21] L.S. Chen. Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction. *PhD thesis, University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering.*, 2000.

[22] Y. Chen, H.D. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K.S. Gopinath, R.W. Briggs, and E.A. Geiser. Using prior shapes in geometric active contours in a variational framework. *IJCV*, 50(3):315–328, 2002.

[23] T. F. Cootes, C.J. Taylor, D. H. Cooper, and J. Graham. Active shape models, their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[24] D. Cremers, T. Kohlberger, and C. Schnorr. Nonlinear shape statistics in mumford-shah based segmentation. *ECCV - LNCS*, pages 93–108, 2002.

[25] D. Cremers, C. Schnorr, and J. Weickert. Diffusion-snakes: Combining statistical shape knowledge and image information in a variational framework. In *In Proceedings of 1st IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pages 137 – 144, 2001.

[26] D. Cremers, N. Sochen, and C. Schnorr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In *in Proc. Int. Conf. Scale Space Theory Computer Vision*, volume 2695, pages 388 – 400, 2003.

[27] L.C De Silva, T. Miyasato, and R. Natatsu. Facial emotion recognition using multimodal information. *In Proc. IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS'97)*, pages 397–401, Singapore, Sept. 1997.

[28] P. Deuflhard. *Newton Methods for nonlinear problems. Affine Invariance and Adaptive Algoritms*, volume 35. Series Computational Mathematics, Springer, 2004.

[29] P. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, Englewood Cliffs, London., www.ph.tn.tudelft.nl/ PRInfo/ books.html www.ph.tn.tudelft.nl/ PRInfo/ books.html www.ph.tn.tudelft.nl/ PRInfo/ books.html, 1982.

[30] R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3d structure recovery of faces. *In Proc. European Conference Computer Vision*, pages 99–113, 2004.

[31] I.L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley Series in Probability and Statistics, 1998.

[32] P. Ekman, W. V. Friesen, and S. Ancoli. Facial signs of emotional experience. *Journal of Personality and Social Psychology.*, 39:1125–1134, 1980.

[33] H El-Ghawalby and E.R Hancock. Geometric characterizations of graphs using heat kernel embeddings. *Mathematics of Surfaces 2009*, (5654):124–142, 2009.

[34] N. I. Fisher. Spherical medians. *J. R. Statist. Soc.B*, 47(2):342–348, 1985.

[35] P.T. Fletcher, S. Joshi, C. Lu, and S.M. Pizer. Principal geodesic analysis for the study of nonlinear statistics of shape. *Transactions on Medical Imaging*, 23:995–1005, 2004.

[36] P.J. Flynn, K.W. Bowyer, and P.J. Philips. Assessment of time dependency in face recognition: An initial study, audioand video-based biometric person authentication. *Audio and Video based biometric person authentication (AVBPA), LNCS 2688*, pages 44–51, 2003.

[37] D. Gavrila. Multi feature hierarchical template matching using distance transforms. In *In Proc of IEEE International Conference on Pattern Recognition*, pages 439– 444, 1998.

[38] J.C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 23:325–328, 1964.

[39] U. Grenander. General pattern theory. *Oxford University*, 1993.

[40] Q. Huaijun and E. R. Hancock. Clustering and embedding using commute time. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(11):1873–1890, 2007.

[41] S. Huckemann and H. Ziezold. Principal component analysis for riemannian manifolds with application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319, 2006.

[42] T. I.T. Jolliffe, I. Principal component analysis. *Springer Series in Statistics*, 2nd edition, NY., 2002.

[43] A.K. Jain, R. P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 2000.

[44] I.T. Jolliffe. Principle component analysis. *Springer-Verlag, New York*, 1986.

[45] J. Jurgen. *Riemannian Geometry and Geometry Analysis.* ISBN 978-3540773405. Springer-Verlag, 5 edition, 2008.

[46] T. Kawano, K. Kato, and K. Yamamoto. A comparison of the gender differentiation capability between facial parts. *In. Proc. of 17th International Conference on Pattern Recognition (ICPR). IEEE*, pages 350–357, 2004.

[47] D. G. Kendall, D. Barden, T.K. Carne, and H Le. *Shape and Shape Theory*. Wiley series in probability and statistics, second edition edition, 1999.

[48] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[49] W. S. Kendall. The diffusion of shape. *Advances in Applied Probability*, 3:428–430, 1977.

[50] W. S. kendall. A survey of the statistical theory of shape. *Statistical Science*, 4(2):111–113, May 1989.

[51] K.I Kim, K. Jung, and H.J. Kim. Face recognition using kernel principal component analysis. *IEEE Processing Letters*, 9(2):40–42, 2002.

[52] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[53] Y. Koren and L. Carmel. Robust linear dimensionality reduction. *IEEE Trans. Visualization and Computer Graphics*, 10(4), 2004.

[54] H. Le and C. G. Small. Multidimensional scaling of simplex shapes. *Pattern Recognition*, 32(9):1601–1613, September 1999.

[55] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:1316, 2000.

[56] Y. Lijun, W. Xiaozhou, S. Sun, W. Jun, and R. J. Matthew. A 3d facial expression database for facial behavior research. *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 211 – 216, 2006.

[57] X. Lu, Y. Chen, and A.K. Jain. Multimodal facial gender and ethinicity identification. *In Proc. Int'l Conf. on Biometrics*, pages 554–561, 2006.

[58] Y. Lu, H. Huang, Y. Chen, and D. Yang. Automatic gender recognition based on pixel-pattern-based texture feature. *J. Real Time Image Processing*, 3:109–116, 2008.

[59] E. Makinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(30):541–547, March 2008.

[60] E. Makinen and R. Raisamo. An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29:1544–1556, 2008.

[61] K. V. Mardia and P. E. Jupp. *Directional Statistics*. John Wiley and Sons Ltd, 2000.

[62] D. Marr. Vision: A computational investigation into the human representation and processing of visual information. *W.H. Freeman*, 1982.

[63] K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions*, (E74(10)):3474–3483, October 1991.

[64] S.J. Maybank. Detection of image structures using the fisher information and the rao metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):1579–1589, 2004.

[65] G.W. Milligan and M.C. Cooper. A study of the comparability of external criteria for hierarchical cluster analysis., 1986.

[66] W. Mio, D. Badlyang, and X. Liu. A computational approach to fisher information geometry with applications to image analysis. *in Energy information geometry with applications to image analysis (EMMCVPR)*, pages 18–33, 2005.

[67] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[68] T. Otsuka and J. Ohya. Recognizing multiple persons' facial expressions using hmm based on automatic extraction of significant frames from image sequences. *In Proc. Int. Conf. on Image Processing (ICIP-97).*, Santa Barbara, CA, USA.:546–549, October 1997.

[69] T.V. Papathomas and A. Gorea. Ambiguity in 3-d patterns induced by lighting assumptions. *Perception*, 19:569–571, 1990.

[70] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine.*, 2(6):559–572, 1901.

[71] E. Pekalska, A. Harol, and R. P.W. Duin. Non-euclidean or non-metric measures can be informative. *SSPR & SPR*, (4109):871–880, 2006.

[72] X. Pennec. Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements. In *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, pages 194–198, 1999.

[73] X. Pennec. Probabilities and statistics of riemannian manifolds: A geometric approach. Research Report RR-5093, INRIA, 2004.

[74] A. Pentland, B. Moghadam, T. Starner, and M. Turk. View-based and modular eigenspaces for face-recognition. In *In Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[75] A. Peter and A. Rangarajan. A new closed-form information metric for shape analysis. *MICCAI, LNCS 4190*, pages 249–256, 2006.

[76] M. W. Rand. Objective criteria for the evaluation of clustering methods, 1971.

[77] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Detection of interest points using symmetry. pages 62–65, Osaka, Japan, Dec. 1990.

[78] A. Robles-Kelly and E.R Hancock. Estimating the surface radiance function from single images. *Graphical Models*, 67(6):518–548, November 2005.

[79] M. Rosenblum, Y. Yacoob, and L.S. Davis. Human expression recognition from motion using a radial basis function network architecture. *IEEE Transactions on Neural Network.*, 7(5):1121–1138, September 1996.

[80] M. Rousson and N. Paragios. Shape priors for level set representations. *ECCV*, pages 78–92, 2002.

[81] Y. Saatci and C.P. Town. Cascaded classification of gender and facial expression using active appearence models. *Proc. International Conference on Automatic face and gesture recognition*, pages 393–400, 2006.

[82] B. Scholkopt, A. Smola, and K.R. Muller. Nonlinear component analysis as a kernel eigenvalues problem. *Neural Computation*, 10(5):1299–1319, 1998.

[83] G. Shakhnarovich, P. Viola, and B. Moghadam. A unified learning framework for real time face detection and classification. *IEEE Int'l Conference Automatic Face and gesture recognition*, pages 14–21, 2002.

[84] E. Sharon and D. Mumford. Two dimensional shape analysis using conformal mapping. *CVPR*, pages 350–357, 2004.

[85] L. Sirovich. Turbulence and the dynamics of coherent structures. *Quarterly of Applied Mathematics*, XLV(3):561–590, 1987.

[86] C. G. Small and H. Le. The statistical analysis of dynamic curves and sections. *Pattern Recognition*, 25:1597–1609, 2002.

[87] C.G. Small. *The Statistical Theory of Shape*. Springer, 1996.

[88] W. Smith and E.R. Hancock. Recovering facial shape using a statistical model of surface normal direction. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1914–1930. Published by the IEEE Computer Society, December 2006.

[89] W. Smith and E.R. Hancock. A new framework for grayscale and colour non-lambertian shape-from-shading. *In Proc. Asian Conference Computer Vision*, pages 869–880, 2007.

[90] W. Smith and E.R. Hancock. Facial shape from shading and recognition using principal geodesic analysis and robust statistic. *Int'l J. Computer Vision*, 1(76):71–91, 2008.

[91] A. Srivastava, S. Joshi, W. Mio, and X. Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):590–602, 2005.

[92] A. Srivastava and W. Mio. Some approaches to shape analysis. 2004.

[93] M. A. Stephens. Use of the von mises distribution to analyse continuous proportions, 1982.

[94] Z. Sun, G. Bebis, X. Yuan, and S.J. Louis. Genetic feature subset selection for gender classification: a comparison study. *WACV*, pages 165–170, 2002.

[95] D.L. Swets and L. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18:831–836, 1996.

[96] D. Terzopoulos and R. Szeliski. Tracking with kalman snakes. *Active Vision*, pages 3–20, 1992.

[97] D. Thomson. *On Growth and Form*. Cambridge University, 1942.

[98] M.E. Tipping and C.M. Bishop. Mixtures of probabilistic principal component analysis. *Neural Computation*, 11(2):443–482, 1999.

[99] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *Neural Computation*, 61(3):611–622, 1999.

[100] A. S. Tolba, A. H. El-Baz, and A. A. El-Harby. Face recognition: A literature review. *International Journal of Signal Processing*, 2(2):88–103, 2005.

[101] W.S. Torgerson. Multidimensional scaling i: theory and method. *Psychometrika*, 17:401–419, 1952.

[102] A. Torsello and E.R. Hancock. Graph embedding using tree edit-union. *Pattern Recogn.*, 40(5):1393–1405, 2007.

[103] G.T Toussaint. Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining., 2005.

[104] N. Troje and H.H Bulthoff. Face recognition under varying poses. the role of texture and shape. *Vision Research*, 36:1761–1771, 1996.

[105] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets,. 22(2):137–154, 2003.

[106] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition*, pages 586–591, June 2001.

[107] Y. Wang and L. Staib. Boundary finding with correspondence using statistical shape models. *In IEEE Conf. Comp. Vision and Patt. Recog.*, pages 338 – 345, 1998.

[108] T. Wilhelm, H.J. Bohme, and H.M Gross. Classification of face images for gender, age, facial expression, and identity. *ICANN*, LNCS 3696:569–574, 2005.

[109] M William, Eysenck, and M. T. Keane. *Cognitive Psychology: a student's handbook.* Fifth Edition, Psychology Press, 2005.

[110] R. C. Wilson, E.R. Hancock, E. Pekalska, and R. P.W. Duin. Spherical embeddings for non-euclidean dissimilarities. *CVPR*, pages 1903–1910, 2010.

[111] P.L. Worthington and E.R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(12):1250–1267, 1999.

[112] Jing Wu. Statistical approaches to gender classification in the surface normal domain. *Thesis*, 2009.

[113] Jing Wu, W.A.P. Smith, and E.R. Hancock. Facial gender classification using shape-from-shading. *Image and Vision Computing*, 20:1039–1048, 2010.

[114] Bai X and E.R Hancock. Heat kernels, manifolds and graph embedding. *SSPR*, (3138):198–206, 2004.

[115] Weiping Xu, E.R. Hancock, and R. C. Wilson. Regularising the ricci flow embedding. *SSPR & SPR*, pages 579–588, 2010.

[116] Weiping Xu, E.R. Hancock, and R.C. Wilson. ectifying non-euclidean similarity data using ricci flow embedding. *ICPR*, pages 3324–3327, 2010.

[117] Z. Xu, L. Lu, and P. Shi. A hybrid approach to gender classification from face images. In IEEE Conference on Computer Vision, editor, *ICPR*, pages 1–4, 2008.

[118] Y. Yacoob and L.S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 18(6):636–642, June 1996.

[119] M.H. Yang and D. Ahuja, N. Kriegman. Face recognition using kernel eigenfaces. *IEEE Int'l Conference Image Processing*, 1:37–40, 2000.

[120] K. Y. Yeung and L. Ruzzo. Principal component analysis for clustering gene expression data. *Oxford University Press*, 17:763–774, 2001.

[121] L. Younes. Computable elastic distances between shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586, 1998.

[122] A. Yuille, D. Cohen, and P. Hallinan. Feature extraction from faces using deformable templates,. In *CVPR*, pages 104–109, 1989.

[123] J. Zhang, Zhang X., H. Krim, and G.G. Walter. Object recognition and recognition in shape spaces. *Pattern Recognition*, 36:1143–1154, 2003.

[124] W. Zhao and R. Chellappa. Sfs based view synthesis for robust face recognition. pages 285–292, 2000.

[125] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surveys*, 35(4):399–458, 2003.

[126] H. Ziezold. On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Trans. 7th Praque Conf. Information Theory, Stat. Dec. Func, Random Processes*, A:491–510, 1977.

List of Symbols

$Exp_x(y)$	The Exponential map of point y at base-point x
$Log_x(y)$	The log map of point y at base-point x
T_xS^2	The tangent plane to a spherical manifold at the point x
μ	The intrinsic mean of a distribution of point on a manifold
u	Log mapped long vector of a needle-map
$d(n_1, n_2)$	Geodesic Distance
$\pi_C(x)$	The projection of a point on a spherical manifold onto a great circle
$\sum$	Covariance Matrix
λ_i, e_i	The ith largest eigenvalues, and the corresponding eigenvector
Φ	The projection matrix formed by the leading eigenvectors
b	Feature vector