Computing gender difference using Fisher-Rao metric from facial surface normals

Simone Regina Ceolin
Centro Universitário Franciscano
UNIFRA
Santa Maria - RS, Brazil
simoneceolin@gmail.com

Edwin R. Hancock*
Computer Science Department
University of York
York, England
erh@cs.york.ac.uk

***Abstract*—The aim in this paper is to explore whether the Fisher-Rao metric can be used to characterise the shape changes due to gender difference. We work using a 2.5D representation based on facial surface normals (or facial needle-maps) for gender classification. 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 low 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 Max Planck and EAR database.**

***Keywords*-Fisher-Rao metric; surface normal; shape-from-shading.**

I. Introduction

Over the past decade there has been a considerable growth in interest in the statistical theory of shape [1],[2]. This field of study has been the result of a synthesis of ideas from a number of different areas including statistics, computer vision, pattern recognition and machine learning, and the realization that the areas share a considerable common ground [3].

Statistical theories of shape variation [4] 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 [1], [2]. 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.

Our aim is to construct a shape-space that can be used to recognise instances of the same face from 2.5D images and also, to construct a model for variations in facial shape due to changes in different gender using information provided by facial needle-maps. A facial needle-map 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 [5]. This 2.5D representation is critical for shape-processing, and can be derived from 2D retinal images using shape-from-shading.

One recent and powerful development in this area has been to explore the use of techniques motivated by information theory, and in particular to use the Fisher-Rao metric to measure the similarities of statistical shape models and construct shape-spaces. In the literature Maybank [6] shows how to use Fisher information for line detection, Mio et al. [7] apply non-parametric Fisher-Rao metrics for image segmentation and Peter [8] has presented a unified framework for shape representation and deformation.

In this paper 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. The reported 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 [4]. 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 of doing this 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 are more directly linked to the physical process of light reflectance.

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 [9] which converts surface normals into a Cartesian form using the equidistance azimuthal projection. With the surface

*Edwin Hancock is supported by a Royal Society Wolfson Research Merit Award and the EU FET project SIMBAD (213250).

normal data in Cartesian form we construct a shape-space for variations due to changes 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. Our goal in this paper is to explore how to use the vMF distribution for shape representation, and in particular to recognise variations in facial shape due to gender difference.

Working in the surface normal domain, 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. In this way facial similarity is measured by the geodesic distance between the shapes on a statistical manifold.

The remainder of the paper is organized as follows. Section II describes how to use the PGA method to construct the statistical model for face needle-maps. Section III describes how the Fisher-Rao metric can be used to measure the similarity of facial needle-maps. Section III-C discusses how multidimensional scaling can be used to embed faces into a low-dimensional pattern space based on the Fisher-Rao metric. Section IV provides some experiments on gender discrimination using Max Planck dataset. Finally, Section VI offers some conclusions and suggests some directions for future research.

II. Cartesian Representation of Vector Fields

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 $U \in S^2(N) = \prod_{i=1}^{N} S^2$.

A. 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 Fig. 1.

The inverse of the exponential map is the *log map*, denoted Log_n. Therefore, the equality $\mathrm{Log}_n(\mathrm{Exp}_n(v)) = v$ 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. $d(n_1, n_2) = \|\mathrm{Log}_{n_1}(n_2)\|$. 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.

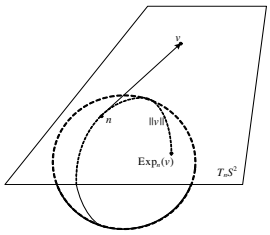


Fig. 1. The exponential map.

B. 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 $\hat{\mathbf{n}}_0 = \frac{\bar{\mathbf{n}}}{|\bar{\mathbf{n}}|}$ where $\bar{\mathbf{n}} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{n}_i$ [10]. 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:

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

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* [11]. This point cannot be found analytically, but can be solved for iteratively using the gradient descent method of Pennec [12]. 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:

$$\mu_{(j+1)} = \mathrm{Exp}_{\mu_{(j)}}\left(\frac{1}{K}\sum_{i=1}^{K}\mathrm{Log}_{\mu_{(j)}}(n_i)\right) \tag{2}$$

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 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:

$$\sigma^2 = \frac{1}{K}\sum_{i=1}^{K} d(\mu, n_i)^2 = \frac{1}{K}\sum_{i=1}^{K} \|\mathrm{Log}_\mu(n_i)\|^2 \qquad (3)$$

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 (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.

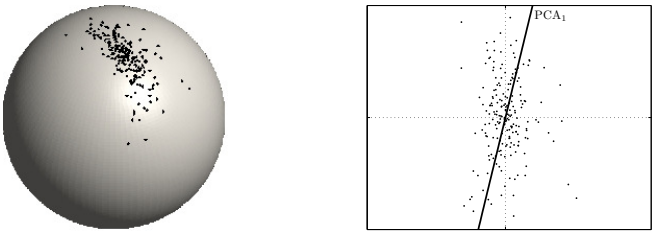
C. Principal Geodesic Analysis for facial Needle-maps

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 $\pi_c : S^2 \to C$ 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 $T_\mu S^2$

$$\log_\mu(\phi_C(n_1)) \approx \sum_{i=1}^{d} v_i . \log_\mu(n_1), \qquad (4)$$

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 2.



(a) (b)

Fig. 2. Projection of surface normals on the unit sphere (a) to points on the tangent plane at the mean (b) [9].

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 [13] 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.

III. Geodesic Distances between Fields of Surface Normals using the Fisher-Rao Metric

The construction of shape-spaces is an emerging and exciting area of statistics, offering many fresh challenges ([14] [15]). The development of a rigorous statistical theory of shape began with the work by Kendall [16] 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 both archaeology and astronomy. Bookstein [17] and Ziezold [18], 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 [19] and [20]. Recent developments in statistical shape theory due to Small ([21]) 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. The aim in this paper is to explore whether 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 (vMF) distribution.In particular we aim to characterise the shape changes due to differences in gender. We make use of the vMF distribution since we are dealing with surface normal data over the sphere $\Re^2$.

A. The von-Mises Fisher distribution (vMF)

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

$$f_p(\underline{x}, \underline{\mu}, \kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \exp(\kappa \underline{\mu}^T \underline{x}) \tag{5}$$

where $\underline{x}$ is a p dimensional vector residing on the hyper-sphere S^{p-1} submersed in $\Re^p$, and $I_l(\kappa)$ 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 room to $f(x|\underline{\mu}, \kappa)$ are concentrated about the mean direction μ. Larger values of κ imply stronger concentration around 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. Mardia et.al [10] give details of the vMF distribution. The distribution is unimodal and rotationally symmetric around the direction $\underline{\mu}$. Finally, the vMF 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 the unit vector $\underline{x}$, i.e., $\underline{x}_1, ...\underline{x}_m$. The estimator of the mean direction is given by

$$\underline{\mu} = \frac{\sum_{i=1}^{m} \underline{x}_i}{||\sum_{i=1}^{m} \underline{x}_i||} \tag{6}$$

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 [22]. It is worth noting that Jupp and Mardia [10] 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.

B. 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 [23],[24],[7].

Let I = [0,1], I for Bessel function and p: $\mathrm{I} \times \Re^k \to \Re^+$, $(\mathrm{x},\underline{\theta}) \mapsto \mathrm{p}(\mathrm{x};\underline{\theta})$, 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:

$$g_{ij}(\underline{\theta}) = \int p(\underline{x}|\underline{\theta}) \frac{\partial}{\partial \theta_i} \log p(\underline{x}|\underline{\theta}) \frac{\partial}{\partial \theta_j} \log p(\underline{x}|\underline{\theta}) d\underline{x}. \tag{7}$$

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 (7) 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 with $\theta = \kappa, \mu_1, \mu_2, \mu_3$.

For simplicity, we concatenate the components of the mean surface normal $\underline{\mu}$ and write $\underline{\theta} = (\kappa, \underline{\mu}^T)^T$ and perform vector-differentiation with respect to μ to simplify our calculations. In the following sections we detail how to compute the element of Fisher Rao information matrix for the von-Mises Fisher distribution.

1) Computing $g_{\kappa,\kappa}$: We commence by computing

$$g_{\kappa,\kappa} = \int_{\underline{x}} f_p(\underline{x}, \kappa, \underline{\mu}) \frac{\partial}{\partial \kappa} \log f_p(\underline{x}, \kappa, \underline{\mu}) \frac{\partial}{\partial \kappa} \log f_p(\underline{x}, \kappa, \underline{\mu}) d\underline{x}. \tag{8}$$

Substituting for the vMF distribution, we have

$$\begin{aligned} g_{\kappa,\kappa} &= \int_{\underline{x}} \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa) e^{-\kappa \underline{\mu}^t \underline{x}}}{\kappa^{\frac{p}{2}-1}} \\ &\quad \left[\frac{\partial}{\partial \kappa} \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} e^{\kappa \underline{\mu}^t \underline{x}} \right]^2 d\underline{x} \end{aligned} \tag{9}$$

Performing the partial derivative with respect to κ, we have:

$$\begin{aligned} g_{\kappa,\kappa} &= \int_{\underline{x}} \frac{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa) e^{-\kappa \underline{\mu}^t \underline{x}}}{\kappa^{\frac{p}{2}-1}} \\ &\quad \left[\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} (\underline{\mu}^t \underline{x}) e^{\kappa \underline{\mu}^t \underline{x}} + \frac{e^{\kappa \underline{\mu}^t \underline{x}}}{(2\pi)^{\frac{p}{2}}} \times \right. \\ &\quad \left(\frac{I_{\frac{p}{2}-1}(\kappa)(\frac{p}{2}-1)\kappa^{\frac{p}{2}-2}}{I_{\frac{p}{2}-1}(\kappa)^2} \right) \\ &\quad - \left. \left(\frac{12\kappa^{\frac{p}{2}-1}(I_{\frac{p}{2}-2}(\kappa)) + I_{\frac{p}{2}(\kappa)}}{I_{\frac{p}{2}-1}(\kappa)^2} \right) \right] d\underline{x} \end{aligned} \tag{10}$$

Moving terms that do not depend on $\underline{x}$ integral, we have:

$$g_{\kappa,\kappa} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}}\frac{1}{I_{\frac{p}{2}-1}(\kappa)}\int_{\underline{x}} e^{\kappa\underline{\mu}^t\underline{x}}\left(\frac{1}{\kappa^2}\right)\left[(\kappa\underline{\mu}^t\underline{x})^2 + 2a(\kappa\underline{\mu}^t\underline{x}) + a^2\right]d\underline{x} \tag{11}$$

$a = \frac{(\frac{p}{2}-1)I_{\frac{p}{2}-1}(\kappa) - \frac{\kappa}{2}\left(I_{\frac{p}{2}-2}(\kappa) + I_{\frac{p}{2}(\kappa)}\right)}{I_{\frac{p}{2}-1}(\kappa)}$
where we can define as:

$$g_{\kappa,\kappa} = \frac{1}{\kappa^2}\left(\kappa^2 < cos\theta_\mu >^2 + 2a\kappa < cos\theta_\mu > + a^2\right) \tag{12}$$

Where, $< cos\theta_\mu >= \int_{\underline{g}} ye^y dy$ when $y = \kappa\mu^T x$
$< cos\theta_\mu >= \int_{\underline{g}} ye^y dy$ when $y = \kappa\mu^T x$
In the above we can set $p = 3$ since we are dealing with a vMF distribution over a 2D field of surface normals.

2) Computing $g_{\kappa,\underline{\mu}}$: We commence by computing

$$g_{\kappa,\underline{\mu}} = \int p(X|\theta)\frac{\partial}{\partial\kappa}\log p(X|\theta)\frac{\partial}{\partial\underline{\mu}}\log p(X|\theta)d\underline{x} \tag{13}$$

Again, substituting for the vMF distribution, we have:

$$g_{\kappa,\underline{\mu}} = \int_{\underline{x}} \frac{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}{\kappa^{\frac{p}{2}-1}}e^{-\kappa\underline{\mu}^t\underline{x}}\left[\frac{\partial}{\partial\kappa}\left(\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}e^{\kappa\underline{\mu}^t\underline{x}}\right)\right]\left[\frac{\partial}{\partial\mu}\left(\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}e^{\kappa\underline{\mu}^t\underline{x}}\right)\right]d\underline{x} \tag{14}$$

Perfoming the partial derivatives with respect to κ and μ and collecting terms together:

$$g_{\kappa,\underline{\mu}} = \left[\frac{2\left(\frac{p}{2}-1\right)I_{\frac{p}{2}-1}(\kappa) - \kappa\left(I_{\frac{p}{2}-2}(\kappa) - I_{\frac{p}{2}(\kappa)}\right)}{2(I_{\frac{p}{2}-1}(\kappa))^2}\right]\int_{\underline{x}}\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}(e^{\kappa\underline{\mu}^t\underline{x}})\underline{x}d\underline{x} \tag{15}$$

Since

$\int_{\underline{x}}\frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}(e^{\kappa\underline{\mu}^t\underline{x}})\underline{x}d\underline{x} = \int_{\underline{x}} f_p(\underline{x},\kappa,\mu)\underline{x}d\underline{x}$

is just the mean $f(\underline{x},\kappa,\mu)$ we have:

$$g_{\kappa,\underline{\mu}} = \left[\frac{2\left(\frac{p}{2}-1\right)I_{\frac{p}{2}-1}(\kappa) - \kappa\left(I_{\frac{p}{2}-2}(\kappa) - I_{\frac{p}{2}(\kappa)}\right)}{2(I_{\frac{p}{2}-1}(\kappa))^2}\right]\underline{\mu} \tag{16}$$

3) Computing $g_{\mu,\mu}$: Finally, we compute:

$$g_{\underline{\mu},\underline{\mu}} = \int_{\underline{x}} p(\underline{x}|\theta)\frac{\partial}{\partial\underline{\mu}}\log p(\underline{x}|\theta)\frac{\partial}{\partial\underline{\mu}}\log p(\underline{x}|\theta)d\underline{x} \tag{17}$$

Again, substituting for the vMF distribution and performing the partial derivative with respect to μ

$$g_{\underline{\mu},\underline{\mu}} = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}\int_{\underline{x}}(e^{-\kappa\underline{\mu}^T\underline{x}})\left[(e^{\kappa\underline{\mu}^t\underline{x}})\kappa\underline{x}\right]^2 d\underline{x} \tag{18}$$

$$\begin{aligned} g_{\underline{\mu},\underline{\mu}} &= \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}}I_{\frac{p}{2}-1}(\kappa)}\int_{\underline{x}}(\kappa\underline{x})^2 e^{\kappa\underline{\mu}^T\underline{x}}d\underline{x} \\ &= \int_{\underline{x}} f_p(\underline{x},\kappa,\mu)(\kappa\underline{x})^2 d\underline{x} = \kappa^2\underline{\mu}\underline{\mu}^T \\ &= \gamma(\kappa)\underline{\mu}\underline{\mu}^T \end{aligned} \tag{19}$$

Substituting for the elements of the metric tensor $g_{\kappa,\kappa}$, $g_{\kappa,\mu}$ and $g_{\mu,\mu}$, the (4×4) Fisher information matrix (M) are:

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

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: $\theta^{a_k} = (\kappa_{a_k}, \underline{\mu}_{a_k})^T$ and $\theta^{b_k} = (\kappa_{b_k}, \underline{\mu}_{b_k})^T$ and the mean concentration parameter and mean vector are $\hat{\kappa} = \frac{1}{2}(\kappa_{a_k} + \kappa_{b_k})$ and $\underline{\hat{\mu}} = \frac{1}{2}(\mu_{a_k} + \mu_{b_k})$.

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

$$\begin{aligned} ds^2_{a_k,b_k} &= \alpha(\hat{\kappa})(\kappa_{a_k} - \kappa_{b_k})^2 \\ &+ 2\beta(\hat{\kappa})\hat{\mu}^T(\kappa_{a_k} - \kappa_{b_k})(\underline{\mu}_{a_k} - \underline{\mu}_{b_k}) \\ &+ \gamma(\hat{\kappa})(\underline{\mu}_{a_k} - \underline{\mu}_{b_k})^T\hat{\mu}\hat{\mu}^T(\underline{\mu}_{a_k} - \underline{\mu}_{b_k}). \end{aligned} \tag{20}$$

Now, to discover the desired geodesic between two parametric densities, we can use the Fisher-Rao metric (7) to calculate the distance between the faces.

$$\begin{aligned} ds^2 &= \alpha(\hat{\kappa})(\hat{\kappa_1} - \hat{\kappa_2})^2 + \underline{\mu}^T(\mu_1 - \mu_2)[2\beta(\hat{\kappa})(\kappa_1 - \kappa_2) \\ &+ \gamma(\hat{\kappa})\mu^T(\mu_1 - \mu_2)] \end{aligned} \tag{21}$$

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

$$D^2_{a,b} = \sum_k ds^2_{a_k,b_k} \tag{22}$$

C. 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) [25]. We compare the results with the other two embeddings systems: heat kernel [26] and commute time [27]. MDS is a family of methods that maps measurements of similarity or dissimilarity among pairs of feature items, into distances between feature points with given coordinates in a low-dimensional space. The first step is to compute the squared distance matrix $DS = [D^2_{a,b)}]a, b = 1, ..., n$. This matrix is subjected to the eigendecomposition $DS = \Phi_D \Lambda \Phi_D^T$ where Λ_D is the diagonal eigenvalue matrix with the eigenvalues ordered in decreasing size along the leading diagonal. The embedding co-ordinate matrix $Y_D = \sqrt{\Lambda_D}\Phi_D^T$ has the vectors of embedding co-ordinates of the n data-points as columns. Both the heat kernel embedding and the commute time embedding commence from the Laplacian matrix. Let $W = \exp[-kDS]$ be the matrix with elements $W(a, b) = \exp[-kDS(a, b)]$ where k is a scaling constant. The Laplacian matrix is $L = D - W$ where D is the diagonal degree matrix with elements $D(a, a) = \sum_{b=1}^{n} W(a, b)$. The eigendecomposition of the Laplacian matrix is $L = \Phi_L \Lambda_L \Phi_L$. From the eigendecomposition it is straightforward to compute the co-ordinate matrices of both the Laplacian eigenmap $Y_L = \sqrt{\Lambda_L}\Phi_L^T$ and the heat-kernel embedding $Y_H = \exp[-\Lambda_L t]\Phi_L^T$. 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 co-ordiante 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. 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 generalization 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 [28],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) recommended the Rand Index as the measure of agreement even when comparing partitions having different numbers of clusters [29],

IV. Experimental Results

Our experiments are concerned with assessing shape variation in fields of surface normals due gender difference. We 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.

We use MDS, heat kernel and commute time embedding technique to analyze the experiments (details about those techniques III-C are described in the Section III-C). 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.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 [29] recommends the Rand Index as the measure of agreement even when comparing partitions having different numbers of clusters.

A. 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 [30] [31] 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 3 and 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 attention 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

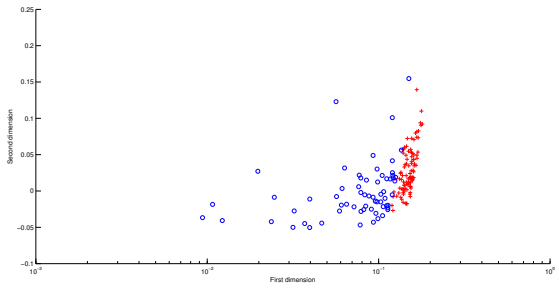


Fig. 3. Gender difference - Max Planck data set.

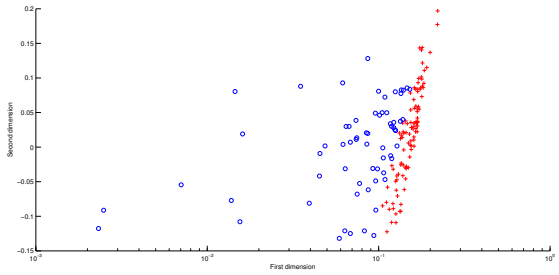


Fig. 4. Gender difference - EAR data set.

of shape alone is a difficult task, and human observers make numerous additional cues such as hair-style.

Table I 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 II, 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 I that the performance from the 1-NN classifier gave the best result using MDS embedding technique for both Data sets. Table II 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 I
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

V. GENDER IDENTIFICATION PERFORMANCE

Compared to the results from Section IV-A, 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

TABLE II
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

TABLE III
GENDER IDENTIFICATION PERFORMANCE

Related Works	**Gender Identification Performance**
Lu Xiaoguang et al. [33]	97%
Wen Yi Zhao et al. [34]	93%
Zing Wu [32]	97%
Ziyi Xu et al. [35]	$92,38\%$
Volker Blanz et al. [36]	$84,75\%$

difference. Also, we can compare our results with the work by Wu [32], 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 are 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. [33] 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. [34] 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 III, we can observe a comparative with the related works. Ziyi Xu et al. [35] proposed a novel hybrid face coding method by fusing appearance features and geometric features. Volker Blanz et al. [36] 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 III, our research compared to related works achieved the best performance. We achieved $97,20\%$ using EAR data set with classification error 1-NN. Also, we have a success using classification error of Rand Index with 88% of identification using Max Plank data set.

VI. Conclusions and Future Work

In this paper we are able to show a notion of distance, using Fisher-Rao metric, between fields of surface normals on a shape manifold. The immediate next step is to construct individual shape-spaces for each class of object. Another line of investigation will be to revisit the problem of computing geodesic distance between needle-maps, in a way that explicitly accounts for the shape of manifold on which they reside.

The overall goal of this paper was to use statistical shape analysis to construct shape-spaces that span 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.

There are clearly a number of ways in which this work may be extended. 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 (Section II) and Fisher-Rao metric (Section III) using 2.5D frontal images from Max-Planck face database [30] and Notre Dame biometric database [37]. 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 incorporating our methods into a set of view-based models similar to those proposed by Pentland et al. [38] and Reisfeld et al. [39] which proposed that a set of separate submanifolds can be obtained by applying PGA to facial needle-maps of each viewpoint.

References

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

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

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

[4] J. Zhang, Z. X., H. Krim, and G. Walter, "Object recognition and recognition in shape spaces," *Pattern Recognition*, vol. 36, pp. 1143–1154, 2003.

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

[6] S. Maybank, "The fisher-rao metric for projective transformations of the line," vol. 63, no. 3, 2005, pp. 191–206.

[7] 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)*, pp. 18–33, 2005.

[8] A. Peter and A. Rangarajan, "Shape analysis using the fisher-rao riemannian metric: unifying shape representation and deformation," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, 2006, pp. 1164–1167.

[9] W. Smith and E. Hancock, "Recovering facial shape using a statistical model of surface normal direction," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12. Published by the IEEE Computer Society, December 2006, pp. 1914–1930.

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

[11] N. I. Fisher, "Spherical medians," *J. R. Statist. Soc.B*, vol. 47, no. 2, pp. 342–348, 1985.

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

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

[14] 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*, vol. 2695, 2003, pp. 388 – 400.

[15] A. Srivastava, S. Joshi, W. Mio, and X. Liu, "Statistical shape analysis: Clustering, learning, and testing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 590–602, 2005.

[16] W. S. Kendall, "The diffusion of shape," *Advances in Applied Probability*, vol. 3, pp. 428–430, 1977.

[17] F. Bookstein, "The measurement of biological shape and shape change," *Lecture Notes in Biomathematics.*, vol. 24, 1978, berlin: Springer.

[18] 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*, vol. A, pp. 491–510, 1977.

[19] D. Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces," *Bulletin of the London Mathematical Society*, vol. 16, no. 2, pp. 81–121, 1984.

[20] A. Torsello and E. Hancock, "Graph embedding using tree edit-union," *Pattern Recogn.*, vol. 40, no. 5, pp. 1393–1405, 2007.

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

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

[23] S. Maybank, "Detection of image structures using the fisher information and the rao metric," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 1579–1589, 2004.

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

[25] H. Le and C. G. Small, "Multidimensional scaling of simplex shapes," *Pattern Recognition*, vol. 32, no. 9, pp. 1601–1613, September 1999.

[26] X. Bai, R. Wilson, and E. R. Hancock, "Manifold embedding of graphs using heat kernel," *Mathematics of Surface XI*, vol. 3604, no. 417, pp. 34–39, 2005.

[27] Q. Huaijun and E. R. Hancock, "Clustering and embedding using commute time," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1873–1890, 2007.

[28] 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*, vol. 33 (4), pp. 593–604, 2005.

[29] *A study of the comparability of external criteria for hierarchical cluster analysis.*, vol. 21, 1986.

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

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

[32] J. Wu, "Statistical approaches to gender classification in the surface normal domain," *Thesis*, 2009.

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

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

[35] Z. Xu, L. Lu, and P. Shi, "A hybrid approach to gender classification from face images," in *ICPR*, I. C. on Computer Vision, Ed., 2008, pp. 1–4.

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

[37] P. Flynn, K. Bowyer, and P. 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*, pp. 44–51, 2003.

[38] 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*, 1994, pp. 84–91.

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