

Gender Classification using Shape from Shading

Jing Wu, W.A.P. Smith, E.R. Hancock

Department of Computer Science, The University of York, York, YO10 5DD, UK.

{jwu, wsmith, erh}@cs.york.ac.uk

Abstract

The aim in this paper is to show how to use the 2.5D facial surface normals (needle-maps) recovered using shape from shading (SFS) to improve the performance of gender classification. We incorporate principal geodesic analysis (PGA) into SFS to guarantee the recovered needle-maps is a possible example defined by a statistical model. Because the recovered facial needle-maps satisfy data-closeness constraint, they not only give the facial shape information, but also combine the image intensity implicitly. Experiments show that this combination gives better gender classification performance than using facial shape or texture information alone.

1 Introduction

Humans are remarkably accurate determining the gender of a subject based on the appearance of the face alone. In fact, an accuracy as good as 96% can be achieved with the hair concealed, facial hair removed and no makeup [6]. In recent years, a lot of effort has been spent on the statistical features based [4], [5], [16], [15], [10] approaches for gender classification. Most of them are based on the 2D intensity information. Although studies [3] have shown that the gender is not only revealed by the 2D face textures, but also has close relationship with the 3D shapes of the human faces, only a few studies have investigated the 3D shapes in gender classification [10]. In [10], Lu and Chen exploit the range information of human faces for gender classification, and propose an integration scheme by combining the registered range and intensity images. The experiment results demonstrate that integrating the 3D range modality provides better classification accuracy than using the 2D intensity modality alone. However, the representation and computation is much more complex for 3D face shapes than 2D face textures. Moreover, due to the immaturity of the 3D sensors in the current market, some typical problems with range images including missing data near dark regions and spikes at the region with high reflectivity would deteriorate the classification.

In this paper, we present a statistical framework for gender classification based on the 2.5D facial needle-maps. The needle-map is a shape representation which can be acquired from 2D intensity images using shape from shading (SFS). Therefore it avoids the problems caused by the immaturity of the current 3D sensors, and provides the shape information from a fixed view point. In [14], a new iterative SFS method is proposed, which not only satisfies the data-closeness constraint (satisfies the image irradiance equation)[17], but also guarantees the projection onto the statistical model that can capture the distribution of the surface normal directions. However, the analysis of the distribution of surface

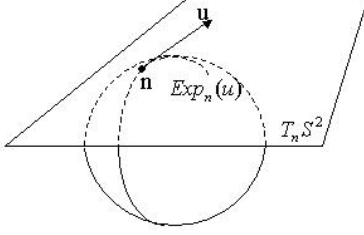


Figure 1: The exponential map.

normals cannot be effected in a linear way, because a linear combination of unit vectors (normals) is not itself a unit vector. Therefore, to construct the statistical model, we make use of principal geodesic analysis (PGA). PGA is a generalization of principle component analysis (PCA). For data residing on a Riemannian manifold, PGA is better suited to the analysis of directional data than PCA.

The idea of our work is to construct a statistical model from a set of ground-truth facial needle-maps, and combine the model and SFS to recover the training and testing needle-maps from the intensity images. Then by performing linear discriminant analysis (LDA) on the PGA parameters of the training needle-maps, we obtain the female and male models which are used to discriminate the genders of the testing faces. Because the SFS method satisfies data-closeness constraint, the recovered facial needle-maps not only represent the 3D shape, but also integrate the 2D texture information. Experiments show that this combination obtains better classification rate than using shape or texture alone.

The outline of the paper is as follows. Section 2 gives the idea of incorporating PGA into SFS to recover facial shapes. Section 3 reviews the LDA method and the probability based classification strategy. In section 4, firstly, a description of the data set and normalization is given, then, the experiment results and discussion are presented.

2 Facial Shape Recovery

We aim to recover the facial needle-maps using an iterative SFS which is augmented by a statistical model that captures variations of the surface normals. The surface normal $n \in R^3$ may be considered as a point lying on a spherical manifold $n \in S^2$, therefore, we turn to the intrinsic mean and PGA proposed by Fletcher et al. [8] to construct the statistical model.

2.1 Principal Geodesic Analysis

If $u \in T_n S^2$ is a vector on the tangent plane to S^2 at n and $u \neq 0$, the exponential map, denoted Exp_n , of u is the point, denoted $Exp_n(u)$, on S^2 along the geodesic in the direction of u at distance $\|u\|$ from n . This is illustrated in Fig. 1. The log map, denoted Log_n is the inverse of the exponential map.

The intrinsic mean is defined as $\mu = \arg \min_{n \in S^2} \sum_{i=1}^N d(n, n_i)$, where $d(n, n_i) = \arccos(n \cdot n_i)$ is the arc length. For a spherical manifold, the intrinsic mean can be found using the

gradient descent method of Pennec [11]. Accordingly, the current estimate $\mu^{(t)}$ is updated as follows: $\mu^{(t+1)} = \text{Exp}_{\mu^{(t)}}\left(\frac{1}{N} \sum_{i=1}^N \text{Log}_{\mu^{(t)}}(n_i)\right)$.

In PGA each principle axis is a geodesic curve. In the spherical case this corresponds to a great circle. For a geodesic G passing through the intrinsic mean μ , π_G may be approximated linearly in the tangent plane $T_\mu S^2$: $\text{Log}_\mu(\pi_G(n_1)) \approx \sum_{i=1}^K V^i \cdot \text{Log}_\mu(n_1)$, where V_1, \dots, V_K is an orthonormal basis for $T_\mu S^2$, which are the principal eigenvectors of the covariance matrix of the long vectors $U = [u^1 | \dots | u^K]$. $u^k = [u_1^k, \dots, u_N^k]^T$ is the log mapped long vector of the k^{th} sample data, in our experiment, the k^{th} training needle-map.

We use the numerically efficient snap-shot method of Sirovich [13] to compute the eigenvectors of L . Accordingly, we construct the matrix $\hat{L} = \frac{1}{K} U^T U$, and find the eigenvalues and eigenvectors. The i^{th} eigenvector e_i of L can be computed from the i^{th} eigenvector \hat{e}_i of \hat{L} using $e_i = U \hat{e}_i$. The i^{th} eigenvalue λ_i of L equals the i^{th} eigenvalue $\hat{\lambda}_i$ of \hat{L} when $i \leq K$. When $i > K$, $\lambda_i = 0$. In our experiments, we use the $K - 1$ leading eigenvectors of L as the columns of the eigenvector matrix (projection matrix) $\Phi = (e_1 | e_2 | \dots | e_{K-1})$, where K is the number of training data.

Given a long vector $u = [u_1, \dots, u_N]^T$, we can get the corresponding PGA parameters $b = \Phi^T u$. Given the PGA parameters $b = [b_1, \dots, b_S]^T$, we can generate a needle-map using: $n_p = \text{Exp}_{\mu_p}((Pb)_p)$.

2.2 Incorporating PGA into SFS

The iterative SFS method used in our experiments satisfies a strict global constraint (projection onto the statistical model) as well as a hard local constraint (satisfaction of the image irradiance equation) in each iteration.

Let $I \in R^N$ denote the intensity image, according to Worthington and Hancock [17], when the surface reflectance follows Lambert's law, 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 $\alpha = \arccos I$. Therefore, the image irradiance equation $I = n \cdot s$ is satisfied by constraining the recovered surface normal to lie on the reflectance cone. While, the recovered surface normal is generated from the best fit PGA parameters, thus the constraint imposed by statistical model is as well satisfied.

So, there are two steps in each iteration (t). First the surface normal at pixel p is generated by: $n_p^{(t)} = \text{Exp}_\mu(u_p^{(t)})$, where $u_p^{(t)} = (\Phi b^{(t)})_p$, and $b^{(t)}$ is the current best fit PGA parameters. Then, to satisfy the data-closeness constraint, each surface normal is rotated back to its closest on-cone position by: $n_p^{(t+1)} = \text{Exp}_s(\arccos(I_p) \frac{\text{Log}_s(n_p^{(t)})}{\|\text{Log}_s(n_p^{(t)})\|})$. This

is demonstrated in Fig. 2 which shows the tangent plane $T_s S^2$. This principal geodesic SFS algorithm can be summarized as follows:

1. Initialize: $N^{(0)} = \mu$, where μ is the intrinsic mean of the model. Set iteration $t = 0$.
2. Estimate PGA parameters: $b^{(t)} = \Phi^T \text{log}_\mu(N^{(t)})$.
3. Update normals: $N^{(t+1)'} = \text{Exp}_\mu(\Phi b^{(t)})$, $N^{(t+1)''} = \text{Exp}_s(\arccos(I) \cdot * \frac{\text{Log}_s(N^{(t)})}{\|\text{Log}_s(N^{(t)})\|})$, where the operator $*$ denotes the component-wise product of any two vectors of the same length. Set $t = t + 1$.
4. Stop if $t > \text{max iterations}$. Set $N' = N^{(t)'}$, $N'' = N^{(t)''}$.

Upon convergence, the output N' is an instance of the statistical model. N'' is the closest on-cone position of N' , therefore it satisfies data-closeness constraint. However,

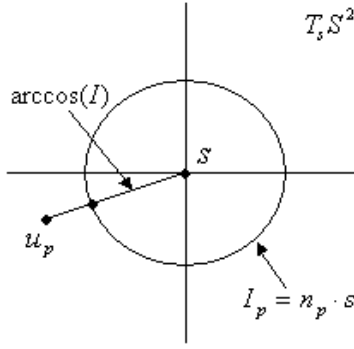


Figure 2: Restoring a normal to the closest position on the reflectance cone.

as stated in [14], the angular difference between N' and N'' is almost solely due to the variation in albedo at the eyes, eye-brows and lips. Therefore, N' can be considered more close to the shape of the face, while N'' combines the facial shape and texture information. Our experiments show the difference between the classification rates of using N' and N'' . The albedo at each pixel can be estimated as well [14]: $\rho_p = \frac{I_p}{s \cdot N'_p}$.

3 Feature Extraction and Classification

The recovered facial needle-maps are represented by the PGA parameters: $b = \Phi^T \log_{\mu}(N)$. However, the parameter vectors inevitably contain information which is either redundant or irrelevant to the gender classification task. Therefore, we extract the most discriminant features for gender using certain discriminating analysis method on the PGA parameters, then the classification is performed on the extracted feature vectors. Linear Discriminant Analysis (LDA) is a simple and widely used discriminating analysis method, and it is especially suitable to the analysis in the PCA subspace. Therefore, in our gender classification task, we choose LDA as the discriminating analysis method.

3.1 Linear Discriminant Analysis

Linear Discriminant Analysis [1] utilizes the class information of the training data to find the vectors in the underlying space that best discriminate among classes. The transform matrix W is chosen in such a way that the ratio of the between class scatter and the within class scatter is maximized. Let the between class scatter be defined as: $S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T$, and the within class scatter be defined as: $S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$, where μ is the mean of all the samples, μ_i is the mean of class X_i , and N_i is the number of samples in class X_i . Then $W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1, w_2, \dots, w_m]$, where $\{w_i | i = 1, 2, \dots, m\}$ is the eigenvectors of $S_W^{-1} S_B$ corresponding to the m largest eigenvalues.

Gender classification is a two-class problem, therefore the class number $c = 2$. The rank of S_w is at most $N - c$, where N is the number of training data. Therefore to avoid

S_w to be singular, the input PGA parameters are truncated to be $N - c$ dimensional by keeping the first $N - c$ components. Because there are at most $c - 1$ nonzero generalized eigenvalues for $S_w^{-1}S_B$, the extracted feature vectors contain only one component in our experiments.

3.2 Classification

After the feature vectors of the training and testing faces have been extracted, we use the a posteriori class probabilities to classify testing faces to one of the genders.

Let $C = \{C_i | i = female, male\}$ denotes the two gender classes, x denotes the feature vector of any testing face. Then according to the Bayes law, the probability that x is of class C_i is:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{\sum_{i=female,male} P(x|C_i)} \quad (1)$$

We assume that the distribution of gender is Gaussian, and the mean and covariance of class C_i is μ_i, σ_i , the a priori probability is $P(C_i) = 0.5$. Then,

$$P(x|C_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right). \quad (2)$$

The a posteriori probabilities $P(C_{female}|x)$ and $P(C_{male}|x)$ are computed by applying (2) and the a priori probabilities into (1). If $P(C_{female}|x) > P(C_{male}|x)$, then the face is classified as female. Otherwise, the face is classified as male.

4 Experiments and Discussion

In this section, we first introduce the data set and the normalization method used in our experiments. Then, we do experiments on the normalized data to evaluate the performance of gender classification based on the recovered facial needle-maps. In experiments, we first examine the performance of SFS. Then the results of gender classification based on the recovered needle-maps are given in comparison with the results based on intensity images. Finally, we compare our method with human classification. The results show that our gender classification strategy is competitive to human classification and achieves better classification rates than using facial shape and texture information alone.

4.1 Data set and normalization

The data set used in our experiments contains 260 3D and corresponding 2D frontal face images (103 females and 157 males) selected from the University of Notre Dame Biometrics Database [9], [7]. To construct the statistical modal and apply the SFS method, the 3D and 2D face images are required to be normalized.

Geometric normalization is needed for the 3D scans which are a set of points $S = \{(x, y, z)\}$. Seven points are manually selected (as shown in Fig. 3): the inside and the outside corners of the left eye (1,2), the inside and the outside corners of the right eye (3,4), the nose tip (5), the middle of the lips (6), the chin point (7). The centers of the

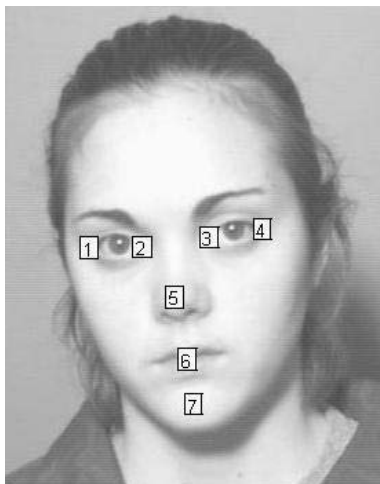


Figure 3: Seven control points.

left and the right eyes, denoted as E_l and E_r , are calculated as the midpoints of 1 and 2, 3 and 4. We first rotate and translate the face scans so that the plane passing E_l , E_r and 7 is perpendicular to the Z -axis, the line passing 1 and 2 is horizontal, and the XY position of 5 is $(0,0)$. After that, we calculate five mean positions (E_l , E_r , 5, 6, 7) for all the 260 scans, and use them as the reference points. Then, we scale the scans to make the distance between 1, 2 identical to the reference. The nose tip 5 gives the centerline for cropping a 100×114 region from the raw 3D scan to create a range image from the depth values. Then, we use the principal warps proposed in [2] to warp the range image so that the XY positions of the five points (E_l , E_r , 5, 6, 7) are identical to those of the reference points. Then linear interpolation is used to fill the holes.

The geometric normalization for the 2D images is almost the same as for the 3D scans, except that the rotation and cropping are performed in XY plane only. Besides the geometric normalization, the 2D images need brightness normalization as well. First, the colored images are converted into greyscale, the intensity at each pixel is the mean value of the three color channels. Then, the intensity contrast is stretched to normalize for ambient lighting variations. Finally, photometric correction and specular subtraction are applied using the method proposed in [12] in order to improve the results of SFS.

After the normalization, the facial needle-maps are calculated from the 3D face scans. Because the cropping step involves re-sampling, some noises are introduced into the needle-maps. Examples of the rendered needle-maps and the normalized 2D images are shown in Fig. 4.

4.2 Experiment Results

In the experiments, 200 normalized face images (needle-maps and corresponding intensity images) are randomly chosen for training, while the remaining 60 are used for testing. In this way, we run the system 6 times and obtain 6 different training and testing data sets. We first apply PGA to the training facial needle-maps to construct the statistical model

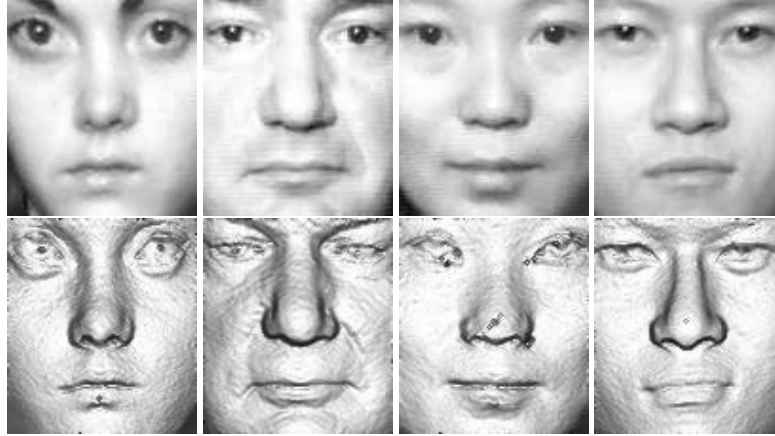


Figure 4: Examples of normalized 2D images (1st row) and surface normals (2nd row).

for SFS. Then, the principal geodesic SFS is used on the training intensity images to get the recovered facial needle-maps for the training data. The recovered needle-maps for the testing data are obtained in the same way. We represent the recovered needle-maps in PGA parameters and apply LDA to extract the feature vectors. The Gaussian models for female and male are constructed from the training feature vectors, and are used to discriminate the gender for the testing faces.

Performance of SFS The recovered needle-maps (N' and N'') and surfaces, together with the estimated albedos of two face images (one male and one female) are shown in Fig. 5. The ground-truth needle-maps and surfaces are also given for comparison. Since N'' satisfy data-closeness, it would appear identical to the image when rendered with a frontal light source. For this reason, we show N'' reilluminated with a light source moved 30° from the viewing direction along the positive x-axis. For comparison convenience, N' and the ground-truth needle-map are also reilluminated in the same way. From the figure, we can see the recovered needle-maps and the surfaces are similar to the ground-truth ones, especially at the cheek and mouth regions in which some gender information is encoded. Both the albedos and N' convey some gender information. However, more gender information is conveyed in N'' which implicitly combines the facial shape and texture.

There are noises in the recovered needle-maps. The reason is that the ground-truth data are noised because of the re-sampling in the alignment, and the statistical model used in SFS is constructed from the noised ground-truth data. Therefore the model captures the noises and introduces it to the recovered needle-maps. Modification of the re-sampling in data alignment to reduce the noise will be needed in our future work.

Gender Classification Results The gender classification rates are shown in Table 1. The experiment results show that both 2D facial texture (intensity images) and 3D facial surface shape (N') shows some capability for gender classification. However, using the recovered facial needle-maps satisfying data-closeness - the implicit combination of

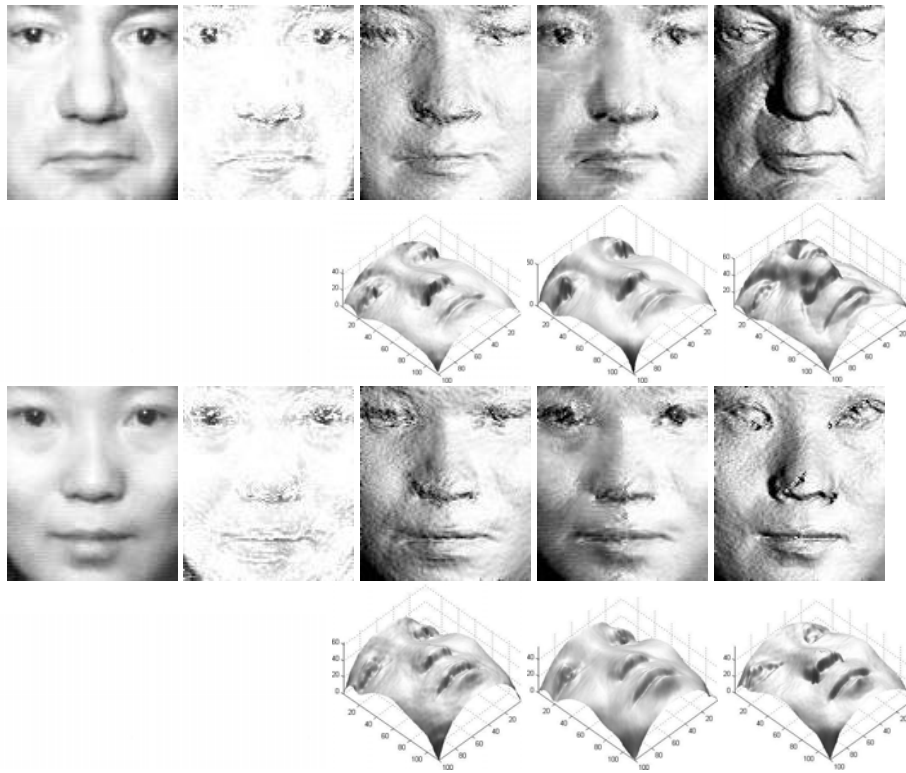


Figure 5: Examples of the results of SFS. From left to right, the columns are: the input intensity images, the estimated albedos, the recovered N' , the recovered N'' , the ground-truth needle-maps. The 2nd and 4th rows are the recovered surfaces of the according needle-maps.

facial texture and shape - achieves better classification rates than using texture or shape individually. Because of the recovery of shape from images, the classification using N' and N'' are more computationally expensive than using the intensity images. However in our experiments the recovery of the shapes for 200 images takes less than 5 minutes.

	Female	Male	Overall
Intensity	83.2% \pm 0.161	89.8% \pm 0.069	87.2% \pm 0.092
N'	68.2% \pm 0.079	70.6% \pm 0.048	69.7% \pm 0.044
N''	95.0% \pm 0.062	92.7% \pm 0.046	93.6% \pm 0.040

Table 1: Gender classification performance.

We use one randomly selected training and testing data set to compare the results of our method using N'' with the results of human classification. We present the 60 testing intensity images to 10 subjects (4 males and 6 females) who are not familiar with the images and are innocent of our research. Their average correct classification rates are calculated. The results are shown in Table 2. From the table, we can see that using the same intensity images (cropped and hair removed), the classification rates achieved using our method are similar to (even better than) the results of human classification.

	Female	Male	Overall
Machine	89.3%	96.9%	93.3%
Human	86.1%	95.6%	91.2%

Table 2: Comparison with human classification.

5 Conclusion

In this paper, we present a gender classification method based on the 2.5D facial needle-maps recovered using SFS. We incorporate PGA into SFS to guarantee the projection onto a statistical model as well as satisfy the data-closeness constraint. The recovered needle-maps implicitly combine the facial shape and texture information. Experiment results show that using the recovered facial needle-maps satisfying data-closeness achieves better gender classification rate than using facial texture or shape information individually.

However, there are some problems that require further investigation. First, modifications of the cropping method in data alignment are needed to reduce the noises caused by re-sampling. Second, the SFS method used is only applicable to grey scale images. The extension to recover needle-maps from color images may improve the gender classification further. Third, apply LDA to the training data is a supervised learning. In the future, we will apply unsupervised learning method to construct models for different genders.

References

- [1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7), 1997.

- [2] F.L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. on PAMI*, 11(6), 1989.
- [3] V. Bruce, A.M. Burton, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, and A. Linney. Sex discrimination: how do we tell the difference between male and female faces? *Perception*, 22:131–152, 1993.
- [4] S. Buchala, N. Davey, R.J. Frank, and T.M. Gale. Dimensionality reduction of face images for gender classification. *Department of Computer Science, The University of Hertfordshire, UK, Technical Report 408*, 2004.
- [5] S. Buchala, N. Davey, T.M. Gale, and R.J. Frank. Principal component analysis of gender, ethnicity, age, and identity of face images. In *IEEE ICMI 2005*, 2005.
- [6] 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.
- [7] K. Chang, K.W. Bowyer, and P.J. Flynn. Face recognition using 2d and 3d facial data. *ACM Workshop on Multimodal User Authentication*, pages 25–32, 2003.
- [8] P.T. Fletcher, S. Joshi, C. Lu, and S.M. Pizer. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. on Medical Imaging*, 23:995–1005, 2004.
- [9] P.J. Flynn, K.W. Bowyer, and P.J. Phillips. Assessment of time dependency in face recognition: An initial study. *Audio and Video-Based Biometric Person Authentication*, pages 44–51, 2003.
- [10] X. Lu, H. Chen, and A.K. Jain. Multimodal facial gender and ethnicity identification. In *ICB 2006*, pages 554–561, 2006.
- [11] X. Pennec. Probabilities and statistics on riemannian manifolds: A geometric approach. *Technical Report RR-5093, INRIA*, 2004.
- [12] A. Robles-Kelly and E.R. Hancock. Estimating the surface radiance function from single images. *Graphical Models*, 67:518–548, 2005.
- [13] L. Sirovich. Turbulence and the dynamics of coherent structures. *Quart. Applied Mathematics*, XLV(3):561–590, 1987.
- [14] W.A.P. Smith and E.R. Hancock. Recovering facial shape and albedo using a statistical model of surface normal direction. *10th IEEE ICCV*, 1:588–595, 2005.
- [15] T. Wilhelm, H.J. Bohme, and H.M. Gross. Classification of face images for gender, age, facial expression, and identity. In *ICANN 2005*, pages 569–574, 2005.
- [16] T. Wilhelm, H.J. Bohme, H.M. Gross, and A. Backhaus. Statistical and neural methods for vision-based analysis of facial expressions and gender. In *IEEE ICSMC 2004*, pages 2203–2208, 2004.
- [17] P.L. Worthington and E.R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. on PAMI*, 21(12):1250–1267, 1999.