Weighted Principal Geodesic Analysis for Facial Gender Classification

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

Department of Computer Science, The University of York, York, YO10 5DD, UK
{jwu,wsmith,erh}@cs.york.ac.uk

Abstract. In this paper, we describe a weighted principal geodesic analysis (WPGA) method to extract features for gender classification based on 2.5D facial surface normals (needle-maps) which can be extracted from 2D intensity images using shape-from-shading (SFS). By incorporating the weight matrix into principal geodesic analysis (PGA), we control the obtained principal axis to be in the direction of the variance on gender information. Experiments show that using WPGA, the leading eigenvectors encode more gender discriminating power than using PGA, and that gender classification based on leading WPGA parameters is more accurate and stable than based on leading PGA parameters.

Keywords: Gender classification, facial surface normals, principal geodesic analysis, weighted principal geodesic analysis, gender discriminating power.

1 Introduction

Humans are remarkably accurate at 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 [1]. In recent years, considerable effort has been spent on the statistical feature based approaches [2], [3], [4], [5], [6] to gender classification. Of these, principal component analysis (PCA) is widely used to reduce the dimensionality of the high dimensional facial data. The aim of PCA is to locate the projections that maximize the variance of the data. However, the projections that maximize the variance usually are not the projections that separate the data into distinct clusters, and so, PCA usually does not reveal cluster structure. Therefore, most of the current approaches employ a second step to extract gender relevant features by performing linear discriminant analysis (LDA) on the PCA parameters. Because of the supervised nature of LDA, this two-step feature extraction strategy is unsuitable for unsupervised learning. Another drawback of the current approaches is that most of them are based on 2D intensity information. Although studies [7] have shown that the gender is not only revealed by 2D facial texture, but also has a close relationship with the 3D shape of the human faces, in fact, only a few studies have investigated the role of 3D shape in gender classification [6]. The computation of 3D face shape representation is significantly more complex

than of 2D face texture. Due to the limited effectiveness and high cost of the 3D sensors in the current market, some typical problems with range images including missing data near dark regions and spikes in regions with high reflectivity would also adversely affect the classification accuracy.

In this paper, we address gender classification using the parameterisation of fields of facial surface normals or needle-maps, and propose a one-step feature extraction method. The needle-map is a 2.5-D shape representation which is intermediate between the 2D intensity image and the 3D surface height function [8]. The representation can be acquired from 2D intensity images using shape-from-shading [9]. It therefore avoids the problems caused by the limitation of current 3D sensors, and is invariant to illumination. To parameterize the facial needle-maps, we use weighted principal geodesic analysis (WPGA). This is a novel variant of principal geodesic analysis (PGA) [10], [11], which constructs a weight matrix making use of the a priori knowledge of the gender discriminating power of different regions of the face, and incorporate the weights into PGA. PGA is a generalization of PCA, for data residing on a Riemannian manifold. As a result, PGA is better suited to the analysis of directional data than PCA. By incorporating weights into the analysis of data, we control the data variance structure so that the variance of gender discriminating regions are larger. In this way, the principal axis obtained after the PGA projection are in the direction of the gender discriminating variance. So, WPGA method improves the projections separating clusters in a manner that is consistent with the projection directions that maximize the variance. Therefore, it enables us to extract the gender discriminating features in a single step.

The outline of the paper is as follows. Section 2 first reviews the concepts of the Log and Exponential maps, spherical medians, and the PGA of needle-maps, and then provides the idea of incorporating weights into PGA. Section 3 states the probability based classification strategy. The details of how to construct the weights and experimental results are presented in Section 4. Finally, Section 5 concludes the paper and offers directions for future investigation.

2   Weighted Principal Geodesic Analysis

The surface normal $n \in R^3$ may be considered as a point lying on a spherical manifold $n \in S^2$, therefore, the intrinsic mean and PGA proposed by Fletcher et al. [10] is suitable to analyze the variations of the surface normals.

2.1   The Log and Exponential Maps

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 exponential and log maps reserve the geodesic distance between two points, i.e. $d(n_1, n_2) = d(u_1, u_2)$, where $u_1 = Log_n n_1, u_2 = Log_n n_2$.
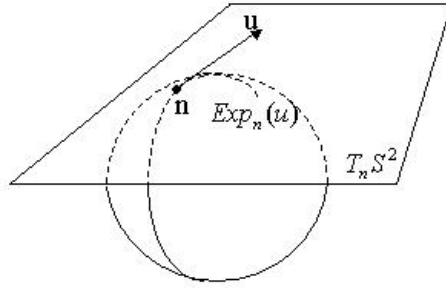
Fig. 1. The exponential map

2.2 Spherical Medians

It is more natural to treat the surface normal as points on a unit sphere: $n_1, \ldots n_N \in S^2$ rather than points in Euclidian space. Instead of the Euclidian mean, we compute the intrinsic mean: $\mu = \arg\min_{n \in S^2} \sum_{i=1}^{N} d(n, n_i)$, 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: $\mu^{(t+1)} = \text{Exp}_{\mu^{(t)}}(\frac{1}{N} \sum_{i=1}^{N} \text{Log}_{\mu^{(t)}}(n_i))$.

2.3 PGA of Needle Maps

PGA is analogous to PCA except that each principal axis in PCA is a straight line, while in PGA each principal axis is a geodesic curve. In the spherical case this corresponds to a great circle. Consider a great circle G on the sphere S^2. To project a point $n_1 \in S^2$ onto a point on G, we use the projection operator $\pi_G : S^2 \longrightarrow G$ given by $\pi_G(n_1) = \text{argmin}_{n \in G}(n_1, n)^2$. 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 $V_1, \ldots V_K$ is an orthonormal basis for $T_\mu S^2$.

Suppose there are K training needle-maps each having N pixel locations, and the surface normal at the pixel location p for the kth training needle-map is n_p^k. We calculate the intrinsic mean μ_p of the distribution of surface normals $n_p^1, \ldots n_p^K$ at each pixel location p. n_p^k is then represented by its log map position $u_p^k = \text{Log}_{\mu_p}(n_p^k)$. $u^k = [u_1^k, \ldots, u_N^k]^T$ is the log mapped long vector of the kth training needle-map. The K long vectors form the data matrix $U = [u^1 | \ldots | u^K]$. The covariance matrix of the data matrix is $L = \frac{1}{K} UU^T$.

The numerically efficient snap-shot method of Sirovich [12] can be used 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 ith eigenvector e_i of L can be computed from the ith eigenvector $\hat{e}_i$ of $\hat{L}$ using $e_i = U \hat{e}_i$. The ith eigenvalue λ_i of L equals the ith eigenvalue $\hat{\lambda}_i$ of $\hat{L}$ when $i \leq K$. When $i > K$, $\lambda_i = 0$. The m leading eigenvectors of L form the projection matrix $\Phi = (e_1 | e_2 | \ldots | e_m)$.

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

2.4   Incorporating Weights into PGA

From above, we can see that PGA, which is a generalization of PCA, captures the directions of the largest variance in the needle-maps. However these variance usually are not associated with the differences in facial shape for different gender. To improve the encoding of gender by the leading eigenvectors, a possible solution is to increase the variance of the normals in the gender discriminating regions, such as eyebrows, nose, etc. Therefore, we introduce the $N \times N$ diagonal weight matrix $W = diag(w_1, \ldots, w_N)$, which gives a weight to each position in the facial needle-map. The positions in the gender discriminating regions are given high weights (w_h), while the other positions are given low weights (w_l). In this way, the normals in the gender discriminating regions have $\frac{w_h}{w_l}$ times larger variance than the normals in the other regions. The leading eigenvectors capture the large variance, and therefore, encode gender discriminating information.

In our experiments, the weight matrix is constructed through the angular difference between the intrinsic means of the female facial needle-maps and male facial needle-maps:

$$w_k = 1 - \exp[-\frac{1}{\sigma^2}(\arccos(\bar{n}_k^m \cdot \bar{n}_k^f))^2] \tag{1}$$

where $\bar{n}_k^m$ is the mean unit surface normal direction for males at the image location where k at. $\bar{n}_k^f$ is the corresponding mean unit normal vector for females. Using the intrinsic means reduces the influence of the differences between identities.

Suppose, $U = [u^1 | \ldots | u^K]$ is the data matrix, where $u^k = [u_1^k, \ldots, u_N^k]^T$ is the log mapped long vector of the kth sample data. The weighted covariance matrix is constructed as $L_W = \frac{1}{K}(WU)(WU)^T$. The snap-shot method of Sirovich are used to compute the eigenvectors of L_W. As stated in [13], 5 gender discriminating significant features will achieve the highest classification rate, we maintain the 5 leading eigenvectors to form the projection matrix $\Phi = (e_1|e_2|e_3|e_4|e_5)$. Given a long vector $u = [u_1, \ldots, u_N]^T$, the corresponding WPGA parameters are computed as $b = \Phi^T(Wu)$.

3   Classification

After the training and test facial needle-maps have been represented by their WPGA parameters, we use the a posteriori class probabilities to classify the test faces to one of the genders.

Let C_f and C_m denote the female and male gender classes, x denote the WPGA parameters of a test facial needle-map. Then according to the Bayes law, the probability that x is of class C_i is:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{\sum_{i \in \{f,m\}} P(x|C_i)} \qquad (2)$$

We assume that the distribution of gender is Gaussian, and that the mean and variance of class C_i are μ_i and σ_i. The a priori probabilities are $P(C_f) = P(C_m) = 1/2$. Then,

$$P(x|C_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{(x - \mu_i)^2}{2\sigma_i^2}). \qquad (3)$$

If $P(C_f|x) > P(C_m|x)$, then the face is classified as female. Otherwise, the face is classified as male.

4   Experiments and Discussion

In this section, we first show how the σ value is determined for the weight matrix construction, and compare the gender discriminating power of the WPGA leading eigenvectors with the PGA leading eigenvectors. Then, the gender classification results obtained using WPGA parameters are compared with those obtained using PGA parameters. The Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany provides the database used in our experiments [14], [15]. This database consists of 200 ground truth facial needle-maps, of which 100 are females and 100 are males. The weight matrix is constructed using all 200 faces. Gender classification is performed by randomly choosing 80 females and 80 males as training data, and using the remaining 40 faces for test. We repeated the randomization 10 times and the classification results are the average of the 10 randomizations.

4.1   Construction and Evaluation of the Weight Matrix

We first examine the selection of the value of σ for the weight matrix construction. Next, we evaluate the performance of WPGA by examining the gender discriminating power of the eigenvectors. The discriminating power is calculated using the criterion function introduced in [16], i.e. $J(\xi) = tr(S_w^{-1}S_b) = \sum_{k=1}^{d} \lambda_k$, where S_w and S_b are the within and between class scatter matrices, and λ_k, $k = 1 \ldots d$ are the eigenvalues of the matrix $S_w^{-1}S_b$.

Determination to the σ value. We construct the weight matrices for 10 different values of σ, and obtain 10 different WPGA projection matrices, from each of which we select the 5 leading eigenvectors. The 10 different sets of WPGA parameters for the 200 faces are obtained accordingly. The gender discriminating power is calculated on the 10 sets of WPGA parameters and are shown in Fig. 2. The σ value is selected to give the largest discriminating power. From Fig. 2, we select σ from the shoulder of the curve of discriminating power and this occurs when $\sigma^2 = 0.11$.
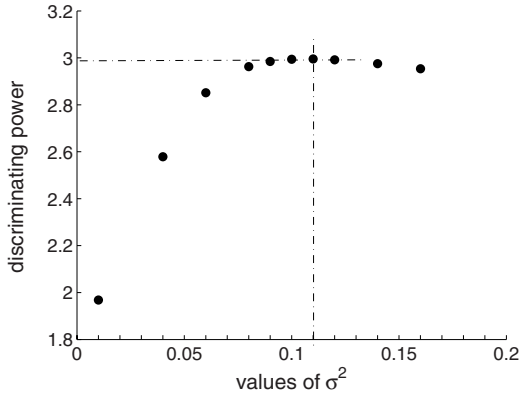
Fig. 2. Selection of σ value

Fig. 3. Construction of the weight matrix. From left to right are the intrinsic mean of female needle-maps, the intrinsic mean of male needle-maps, and the constructed weight matrix.

The intrinsic means of the 100 female needle-maps and the 100 male needle-maps and the weight matrix constructed using $\sigma^2 = 0.11$ are shown in Fig. 3. From the figure, it is clear that the constructed weight matrix has high weights in regions around the eyebrows, nose, and mouth.

Discriminating power. After the construction of the weight matrix, we use it in conjunction with principal geodesic analysis to obtain the projection matrix. The discriminating power of each of the 10 leading WPGA eigenvectors is shown in the left hand panel of Fig. 4, and is compared with those of the 10 leading PGA eigenvectors which are shown in the right hand panel of Fig. 4. From the figure, we see that although the gender discriminating power of the WPGA eigenvectors is not in descending way, it is concentrated in the first 5 eigenvectors. The first 5 WPGA eigenvectors have the 5 largest discriminating power. By comparison, the discriminating power of PGA eigenvectors are more widely distributed. For example, the 8th PGA eigenvector has the 5th largest discriminating power,
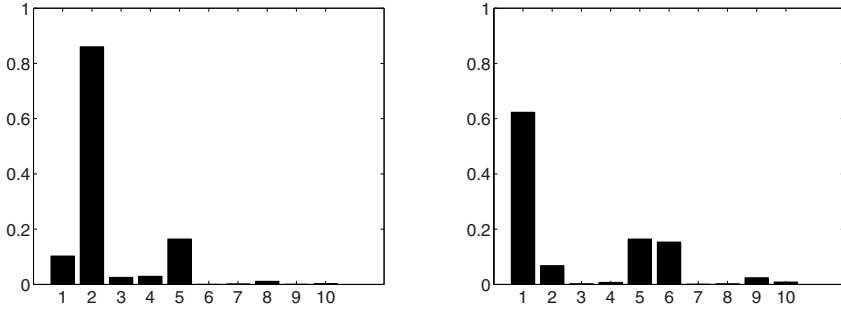
Fig. 4. Discriminating powers of the leading 10 eigenvectors. The left panel is for WPGA and the right panel is for PGA.

and the discriminating power of the 10th PGA eigenvector is larger than that of the 3rd eigenvector. This confirms our assumption that incorporating gender relevant weights into PGA results in better encoding of gender information in the leading eigenvectors. However, in WPGA, the discriminating power of the first and second eigenvectors seem to be swapped, which need further investment in the future.

Fig. 5 shows the plots of cumulative gender discriminating power for the first m ($m = 1 \cdots 20$) WPGA and PGA eigenvectors. From the figure, it is clear that WPGA has significantly larger cumulative discriminating power than PGA when $1 < m < 10$. This gives further confirmation that the first few WPGA eigenvectors encode most of the gender discriminating power, while the
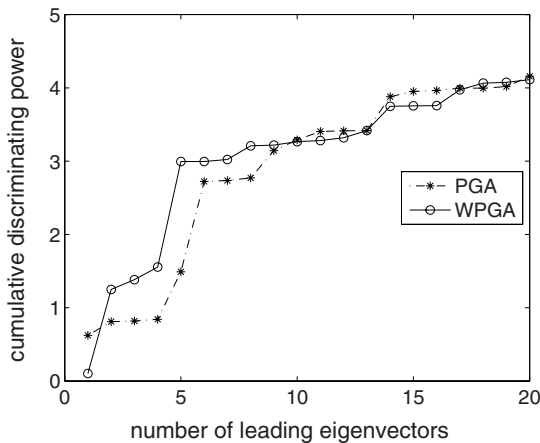

Fig. 5. Cumulative discriminating powers

gender discriminating power encoded in the PGA eigenvectors are more uniformly distributed.

4.2   Classification Results

We randomly selected 80 female needle-maps and 80 male needle-maps from the 200 available for use as training data, and the remaining 40 as test data. We first apply WPGA to the training data to obtain the projection matrix. Then, the construction of separate models for females and males, and gender classification on the test data are both performed on the 5 leading WPGA parameters. We repeat the randomization procedure 10 times. The average classification rates and variance obtained using WPGA and PGA are shown in Table 1, from which it is clear that gender classification performed on the leading WPGA parameters gives improvements not only on the classification rates, but also on the stability of the classification when compared with PCA.

Table 1. Gender classification rates

	Female	Male	Overall
$WPGA$	$93.50\% \pm 0.0450$	$91.00\% \pm 0.0490$	$92.25\% \pm 0.0361$
PGA	$84.00\% \pm 0.1068$	$84.50\% \pm 0.1083$	$84.25\% \pm 0.0448$

5   Conclusion

In this paper, we describe a weighted PGA method to extract gender discriminant features from 2.5D facial needle-maps in a single step, and perform gender classification using the WPGA parameters. Experimental results show that the leading WPGA eigenvectors encode more gender discriminating power than the leading PGA eigenvectors. Moreover, gender classifications based on WPGA parameters achieve more accurate and more stable results than those based on PGA parameters.

There are several potentially interesting avenues for future investigation. First, instead of using ground truth facial needle-maps, we will apply the method to needle-maps recovered from facial images using SFS. Second, we will improve the generalization of the weight matrix. Third, unsupervised learning using the EM algorithm will be used to perform gender classification on the WPGA parameters.

References

1. Burton, A.M., Bruce, V., Dench, N.: What's the Difference Between Men and Women? Evidence from Facial Measurement. Perception 22, 153–176 (1993)
2. Buchala, S., Davey, N., Frank, R.J., Gale, T.M.: Dimensionality reduction of face images for gender classification. Department of Computer Science, The University of Hertfordshire, UK, Technical Report 408 (2004)

3. Buchala, S., Davey, N., Gale, T.M., Frank, R.J.: Principal Component Analysis of Gender, Ethnicity, Age, and Identity of Face Images. In: Proc. IEEE ICMI (2005)
4. Wilhelm, T., Bohme, H.J., Gross, H.M., Backhaus, A.: Statistical and Neural Methods for Vision-based Analysis of Facial Expressions and Gender. In: SMC 2004, pp. 2203–2208. IEEE omnipress (2004)
5. Wilhelm, T., Bohme, H.J., Gross, H.M.: Classification of Face Images for Gender, Age, Facial Expression, and Identity. In: Duch, W., Kacprzyk, J., Oja, E., Zadrożny, S. (eds.) ICANN 2005. LNCS, vol. 3696, pp. 569–574. Springer, Heidelberg (2005)
6. Lu, X., Chen, H., Jain, A.K.: Multimodal Facial Gender and Ethnicity Identification. In: ICB, HongKong, pp. 554–561 (2006)
7. Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., Linney, A.: Sex discrimination: how do we tell the difference between male and female faces? Perception 22, 131–152 (1993)
8. Marr, D.: Vision. W.H. Freeman, San Francisco (1982)
9. Smith, W.A.P., Hancock, E.R.: Recovering Facial Shape and Albedo using a Statistical Model of Surface Normal Direction. Tenth IEEE International Conference on Computer Vision 1, 588–595 (2005)
10. Fletcher, P.T., Joshi, S., Lu, C., Pizer, S.M.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging 23, 995–1005 (2004)
11. Pennec, X.: Probabilities and statistics on riemannian manifolds: A geometric approach. Technical Report RR-5093, INRIA (2004)
12. Sirovich, L.: Turbulence and the dynamics of coherent structures. Quart. Applied Mathematics XLV(3), 561–590 (1987)
13. Wu, J., Smith, W.A.P., Hancock, E.R.: Learning Mixture Models for Gender Classification based on Facial Surface Normals. In: IbPRIA 2007, vol. 4477, pp. 39–46 (2007)
14. Troje, N., Bulthoff, H.H.: Face recognition under varying poses: The role of texture and shape. Vision Research 36, 1761–1771 (1996)
15. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces. In: SIGGRAPH 1999 Conference Proceedings, pp. 187–194 (1999)
16. Devijver, P., Kittler, J.: Pattern Recognition: A Statistical Approach. Prentice Hall, Englewood Cliffs (1982)