

Biologically significant facial landmarks: How significant are they for gender classification?

Syed Zulqarnain Gilani, Faisal Shafait and Ajmal Mian
School of Computer Science and Software Engineering
The University of Western Australia
Email: zulqarnain.gilani@uwa.edu.au

Abstract—Automatic gender classification has many applications in human computer interaction. However, to determine the gender of an unseen face is challenging because of the diversity and variations in the human face. In this paper, we explore the importance of biologically significant facial landmarks for gender classification and propose a fully automatic gender classification algorithm. We extract 3D Euclidean and Geodesic distances between these landmarks and use feature selection to determine the relative importance of the biological landmarks for classifying gender. Unlike existing techniques, our algorithm is fully automatic since all landmarks are automatically detected. Experiments on one of the largest 3D face databases FRGC v2 show that our algorithm outperforms all existing techniques by a significant margin.

Keywords: Gender classification, 3D face, biologically significant landmarks, geodesic distances, feature selection.

I. INTRODUCTION

Facial gender classification has applications in human-computer interaction, surveillance, content-based retrieval, biometrics, demographic studies and targeted advertising [1]. It can also be used as a preprocessing step for person identification based on face biometrics. However, facial gender classification has received less attention in the computer vision literature compared to face recognition. To determine the gender of an *unseen face* is a challenging task for computers due to the diversity and variations in the human face. Although, this task is considered trivial for the human visual system, misclassifications are still possible in the absence of external (not specific to the face region) cues such as hair, cosmetics and texture. Burton et al. [2] show that humans can perform this activity with an accuracy of 96% in the absence of external cues while Bruce et al. [3] claim that on an average humans can make decisions on the gender of faces in 613 to 620 milliseconds. Current computer based facial gender classification algorithms are far behind the accuracy of human visual perception.

The search for discriminatory facial features for gender classification has been a topic of interest for more than two decades. Psychologists have experimented on the human visual system to identify the features that it looks for while classifying the sex of a face. Enlow et al. [4] argue that men have a wider and longer nose as compared to women and that the male forehead is more curved and sloping than the female forehead. Similarly Shepherd et al. [5] claim that female nose is less pointed than the male nose. Farkas et al. [6] identified a set of 23 biologically relevant anatomical facial landmarks (see Figure. 2a.) to extract a set of Euclidean distances, ratios and angles for facial anthropometry. Since gender is

a biological trait, it makes sense to focus on biologically significant landmarks for facial gender classification. However, as our literature survey will show, current computer vision techniques for facial gender classification have paid little attention to the biological significance of features that are used for classification.

Burton et al. [2] manually annotated 73 landmarks on a private dataset of 91 male and 88 female faces. The majority of these landmarks had biological significance and 2628 distances could be extracted from them. However, due to limited computational capability, the authors hand picked only 19 two and three dimensional distances and their corresponding ratios that are discovered most discriminatory by psychological studies for gender classification. They argued that a subset of 16 measurements can classify facial gender with 94% accuracy. Han et al. [7] employed more complex features to classify gender on a small public face dataset of only 61 subjects. They used the SVM classifier on normalized areas and volumes of five local craniofacial regions that are critical to gender distinction, i.e. forehead, eyes, nose, mouth and cheeks. Five landmarks were annotated to identify each region which was then manually extracted. Using 5-fold cross validation, they reported 82.6% accuracy. Lu et al. [8] used a semi-automated method for gender classification on the University of Notre Dame (UND) and Michigan State University (MSU) face databases containing 1240 3D scans of 376 subjects. The authors manually annotated six landmarks on each face to normalize them and then used a 10x8 grid on the middle face to extract feature vectors. Gender classification was performed by training two SVMs directly on the range and intensity values of the extracted grid. The authors reported 85.4% classification

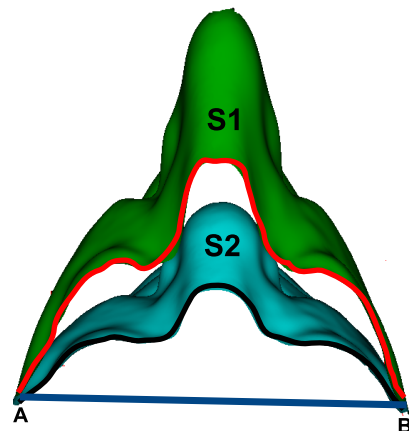


Fig. 1. Sliced cross sectional facial view generated from two models. The Euclidean distance (blue line) between points A and B on both surfaces is the same. However, the geodesic distance between these points on face S1 (red line) is larger than that on S2 (black line).

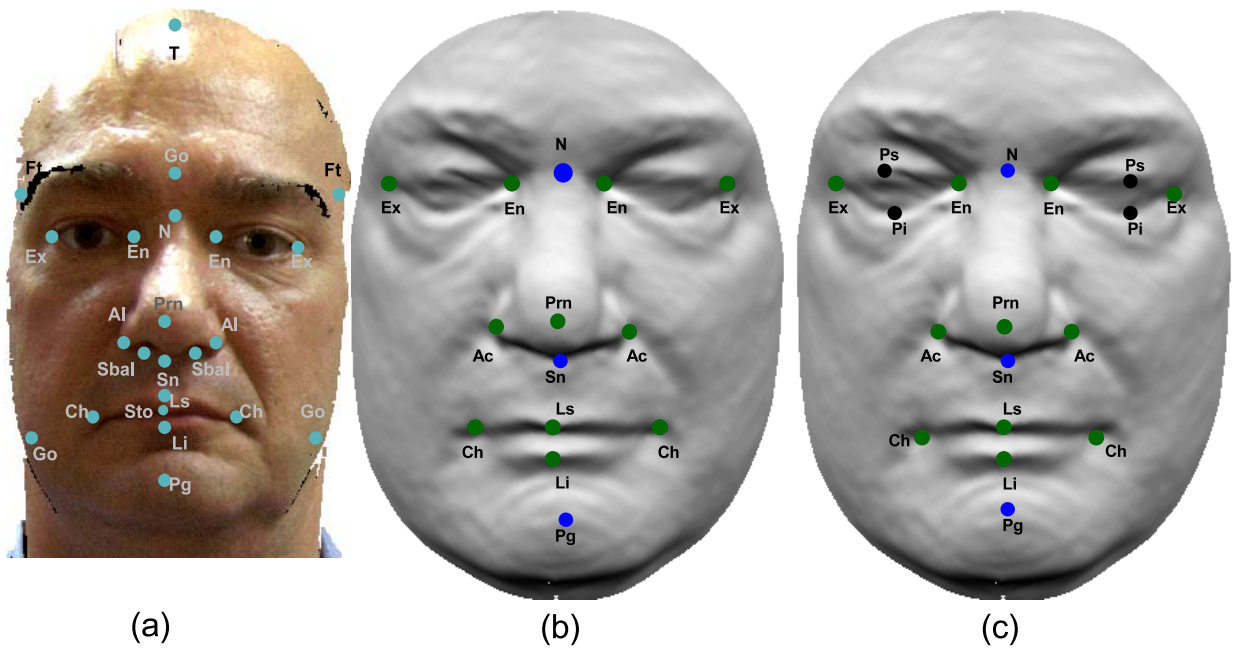


Fig. 2. Landmarks on a 3D face. (a) 23 biologically significant anatomical facial landmarks [6]. (b) 15 MLB landmarks (c) 14 MGT landmarks. Green coloured points are the common landmarks between MGT and MLB while blue points on MLB are the three landmarks later replicated on MGT landmarks to bring the total to 18. See section II and IV-B for details.

accuracy for shape data.

Recently Ballihi et al. [9] performed fully automatic gender classification using the Face Recognition Grand Challenge (FRGC) v2 database. The authors used 3D geometric features to classify gender in 466 unique subjects of the dataset. They used SVM and Neural Networks on radial and circular curves as features extracted from the 3D mesh using the nosetip as the centre. Although these curves represent the shape yet they were not extracted from biologically significant anatomical facial landmarks. Perhaps, this is why the authors achieved an accuracy of only 86.05% using 10 fold cross validation. Similarly a region based fully automated gender classification algorithm was proposed by Hu et al. [10] using 15 automatically detected biologically significant facial landmarks which were used as features. These landmarks were then grouped into five gender discriminating regions of internal face, upper face, left eye, lower face and nose regions. Separate SVMs were trained on the features and the results were fused. They report 94.3% accuracy, however, their database is private, consists of only 216 subjects of the same ethnicity and has no expression variations.

Based on our literature survey, no current technique simultaneously satisfies the following requirements (1) full automation (2) high accuracy (3) use of biologically significant landmarks or features (4) validation on a large public dataset containing multiple ethnicities and facial expressions. The use of biologically significant landmarks is not only expected to give higher facial gender classification accuracy, interpretation of the results will be much easier from a biological point of view. Further details about the motivation for using these landmarks are given in Section III-B.

In this work, we propose a fully automatic facial gender classification algorithm using 3D Euclidean and Geodesic

distances between biologically significant landmarks. Unlike most existing techniques, the landmarks are also automatically detected. Studies suggest that geodesic distances may represent 3D models in a better way compared to 3D Euclidean distances [11]. Figure. 1 shows two face models with similar Euclidean distance between points A and B but different geodesic distances. It can also be argued that geodesic distances are more robust to changes in facial expressions [12] unless discontinuities, like open mouth, are introduced.

We use minimum Redundancy maximum Relevance algorithm [13] combined with a forward selection wrapper targeting a linear classification algorithm to select a subset of the hundreds of possible Euclidean and Geodesic distances. Using 10-fold cross validation on one of the largest public 3D face databases (FRGC v2) [14], we achieve 96.12% accuracy. Separate experiments on the neutral and non-neutral faces of the database show that our algorithm is robust to facial expressions. Comparison of gender classification accuracy using automatic and manually annotated biological landmarks shows only 2% gain in accuracy. Comparison with Ballihi et al. [9] in a similar experimental setup shows that our algorithm achieves 97% accuracy compared to the 86% accuracy achieved the former. We also provide an insight into the relative significance of biological landmarks for facial gender classification.

II. 3D FACE DATA AND LANDMARKS

We have used the public FRGC v2 [14] database consisting of 4007 3D faces of 466 subjects obtained with the Minolta scanner. The database contains significant variations in age, expressions, pose and ethnicity. To extract facial features we have used the 14 automatically detected biologically relevant landmarks from Creusot et al. [15]. The authors have proposed a machine learning approach to detect these facial landmarks.

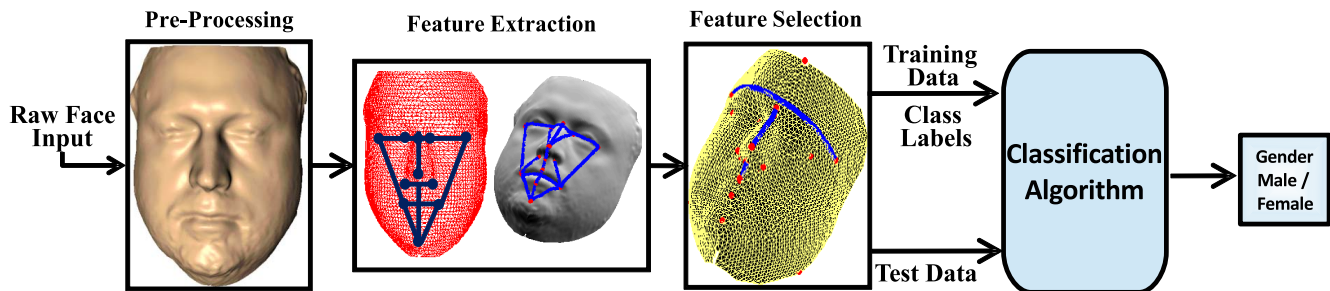


Fig. 3. Block diagram of the proposed gender classification algorithm.

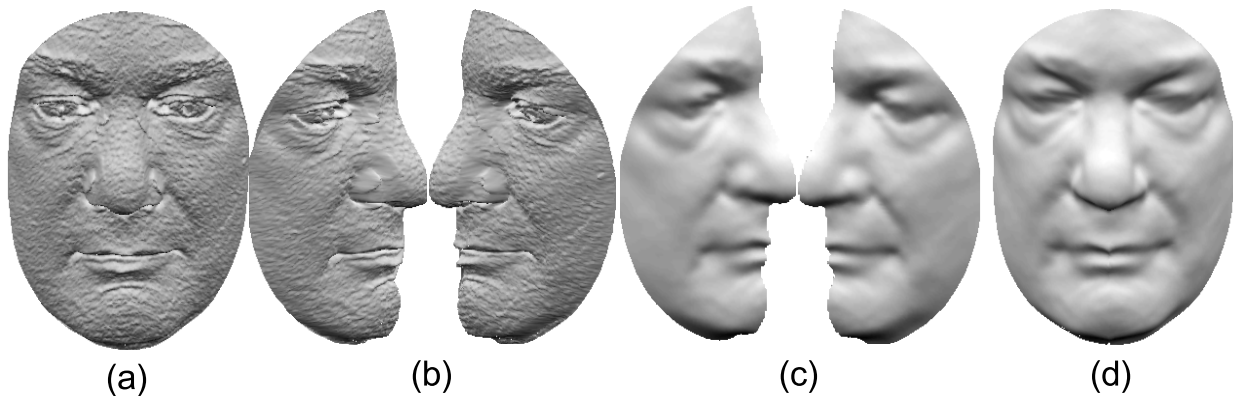


Fig. 4. Different steps in preprocessing. (a) The raw input face (b) Bisected raw face rotated by 45° . Notice the irregular triangulation near the alare curvature due to self occlusion (c) Processed face (d) Processed face stitched back seamlessly.

They train an off line LDA classifier on a set of 200 faces and a landmark model using a myriad of local descriptors for each landmark. Each landmark detection is then reduced to a two class classification problem and the final result is fused to obtain 14 landmarks. The landmarks are shown in Figure. 2b and will be referred to as MLB (Machine Learning Based) landmarks owing to the technique used to detect them. We also extract features from 15 manually annotated landmarks [16] (see Figure. 2c) in order to validate the accuracy of automatically detected landmarks. This set of landmarks will be referred to as MGT (Manual Ground Truth). Both datasets are devoid of the Gonion (Go) and Frontotemporale (Ft), while only the MGT set of landmarks does not contain the Nasion (N) and Subnasale (Sn) which, as our results will show later, are required for deriving critical gender differentiating features like jaw width (Go-Go), forehead width (Ft-Ft), nasal bridge length (N-Prn) and upper lip height (Sn-Sto). Notice that the Gonions (Go) are palpable soft tissue landmarks while Frontotemporales (Ft) do not have a distinguished shape and as such these landmarks are difficult to detect automatically. Details of the dataset and the landmarks are given in Table I.

III. PROPOSED ALGORITHM

An overview of our gender classification algorithm is given in Figure. 3 and its different components are explained below.

A. Pre-processing

The pose of each 3D face is corrected to a canonical form based on four landmarks (Ex(L), Ex(R), N and Prn). This step

TABLE I. DETAILS OF FRGC VER2 DATASET.

Attributes	Details
Total Scans / Persons	4007 / 466
Male/Female	264 / 202
Age (Mean \pm SD)	17-69 yrs (22.54 \pm 6.54)
Expressions	Multiple
Main Ethnicities	White, Asian, Hispanic, Black
Pose Variation	$\pm 10^\circ$
Landmarks Available	
MLB-14: 14 Automatically detected [15]	
MGT-15: 15 Manually annotated ground truth [16]	
MLB-11, MGT-11: 11 landmarks common between MLB-14 & MGT-15	
MGT-18: MGT after adding 3 landmarks from MLB-14 to MGT-15	
MLB-14, MGT-14: 14 landmarks common between MLB-14 & MGT-18	

is required to eliminate any error due to pose in the extraction of geodesic distances. Holes are filled and noise removed by re-sampling the 3D face on a uniform grid using gridfit [17] algorithm which estimates a surface on a 2D grid based on the depth data. The algorithm extrapolates the data to the grid boundaries and uses a modified ridge estimator to generate the surface. Since some portions of the face are expected to be self occluded (e.g. region around Ac) when re-sampled on a grid, we bisect the 3D face at the nose tip and rotate each half by 45° before re-sampling to mitigate this problem. Besides hole filling, another advantage of bisecting and rotating the halves before resampling is that the resulting 3D face has a more uniform sampling in the 3D space. The processed halves are then rotated back and stitched seamlessly to form a single mesh. Figure.4 shows the different steps in preprocessing.

B. Feature Extraction

In this paper we use two feature types, i.e. 3D Euclidean and the geodesic distances between the biologically relevant landmarks. The motivation for using these landmarks and Euclidean distances comes from the fact that they represent the sexual dimorphism of the face [18]. In clinical practice they are used to measure a quantitative dimension for the morphological deviation from the normal face [19]. These features have also been used to delineate syndromes by analysing the subtle difference in the facial distance measures between syndrome affected subjects and control [20]. Psychologists have used these features to measure the amount of facial masculinity/femininity [21]. We have selected the facial landmarks that relate to the bony structure of the face which is affected by the ratio of testosterone to estrogen during adolescence [22]. Being primarily a male sex hormone, high ratio of testosterone in men influences facial bone growth [23]. It is also believed that facial masculinity is associated with levels of circulating testosterone in men [24]. Hence it is intuitive to use features extracted from these bony landmarks for facial gender classification. Given L landmarks on a 3D face, we extract $\binom{L}{2}$ combinations of 3D Euclidean and geodesic distances between them. Extraction of Euclidean distances is a straight forward task. We define geodesic distance as the length of the curve generated by orthogonal projection of the Euclidean line on the 3D facial surface. This is precisely the reason for normalising the pose of each 3D face as variation in pose can present a different surface to the viewing angle. Less curved distances like upper lip height (Sn-Sto) are modelled by a second order polynomial while more curved distances with multiple inflection points, like the biocular width (Ex-Ex) are modelled by higher order polynomials. Figure. 5 shows some of the 3D Euclidean and geodesic features extracted using the MGT and MLB sets of landmarks.

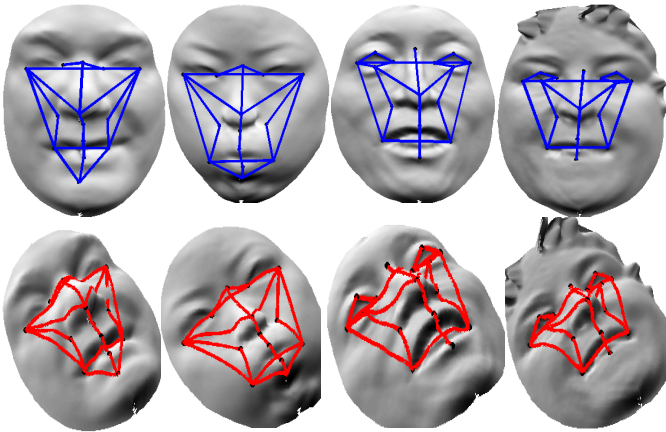


Fig. 5. Examples of some of the 3D Euclidean (top row) and geodesic (bottom row) distances extracted using MGT (first two columns) and MLB (last two columns) landmarks.

C. Feature Selection

Feature selection is performed using the minimal redundancy maximal relevance (mRMR) algorithm packed in a forward-selection wrapper [13]. The algorithm first calculates the intrinsic information (relevance) within a feature and also

the mutual information (redundancy) among the features to segregate different classes. Then it maximizes the relevance and minimizes the redundancy simultaneously. Let $X \in \mathbb{R}^{m \times n}$ be the feature matrix with m observations and n features, F be the target reduced feature set and c be any arbitrary class from the set of classes C , then relevance is defined by,

$$D(F, c) = \frac{1}{|F|} \sum_{x_i \in F} I(x_i; c), \quad (1)$$

and redundancy is defined by,

$$R(F) = \frac{1}{|F|^2} \sum_{x_i, x_j \in F} I(x_i; x_j), \quad (2)$$

where $I(x; y)$ is the mutual information between x and y . Maximal relevance and minimal redundancy is obtained by taking the maximum and minimum values of (1) and (2) respectively. The goal of simultaneously maximizing the relevance and minimising the redundancy is achieved by maximizing the function $\Gamma(D, R)$ where,

$$\Gamma(D, R) = D - R, \quad (3)$$

or

$$\Gamma(D, R) = \frac{D}{R}, \quad (4)$$

where equation (3) is the Mutual Information Difference and equation (4) is the Mutual Information Quotient formulation of mRMR algorithm. Since our feature set is small we find the classification accuracy yielded by both formulations and use only the one giving the maximum accuracy on training data. The reduced number of candidate features k is selected by first obtaining n feature sets F_n using the mRMR sequential search (Eq. 3 or 4 depending on which one gives better accuracy). More specifically $F_1 \subset F_2 \subset \dots \subset F_{n-1} \subset F_n$. Next we compare the classification accuracy for all feature subsets $F_1, \dots, F_k, \dots, F_n$ ($1 < k < n$) to find a range for k where the classification accuracy is maximum. Finally, we select a compact set of features by exploiting the forward-selection wrapper [25]. The wrapper first searches for a single feature Θ_1 from the feature set F_k which gives maximum classification accuracy. Then, from the subset $\{F_k - \Theta_1\}$ we search for another feature such that the subset $\{\Theta_1, \Theta_2\}$ gives the maximum accuracy irrespective of the previous one. This is a deviation from the original mRMR algorithm [13] which desires a feature subset that produces better or equal accuracy than the previous subset in order to minimize the number of evaluations due to greater number of candidate features in F_k . Since our original feature set X contains less than 310 features and the size of candidate feature set F_k is even smaller than X , therefore, we let the wrapper evaluate all possible subsets of F_k in a forward selection scheme enabling us to find the reduced feature subset that gives the best accuracy. Consequently, we obtain a feature set $\{\Theta_1, \dots, \Theta_p, \dots, \Theta_k\}$ where $1 < p < k$ and we select the feature subset $\{\Theta_1, \Theta_2, \dots, \Theta_p\}$ which corresponds to the highest accuracy. Note that this is the most compact feature subset as $1 < p < k < n$. Figure. 6 denotes the number of features selected for 14 MLB and 15 MGT landmarks by the feature selection algorithm.

D. Gender Classification

Next we train an LDA classifier using an exclusive set of training data. Let $\mathbf{X}_i \in \mathbb{R}^{p \times n_i}$ be the matrix of features of class i with n_i samples and p features as chosen in the feature selection step. LDA maximizes the ratio of *between-class scatter* to *within-class scatter*. Between-class scatter is defined as,

$$\mathbf{S}_B = \sum_{i=1}^c n_i (\mu_i - \mu) (\mu_i - \mu)^\top, \quad (5)$$

and within-class scatter is defined as

$$\mathbf{S}_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i) (x_k - \mu_i)^\top, \quad (6)$$

where μ is the mean of all classes, μ_i is the mean of class \mathbf{X}_i and n_i is the number of samples in \mathbf{X}_i . Fisher [26] proposed to maximise the ratio between \mathbf{S}_B and \mathbf{S}_W relative to the projection direction by solving

$$J(w) = \arg \max_w \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}}. \quad (7)$$

By differentiating the equation with respect to \mathbf{w} and equating it to zero, we get $\mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w} - J \mathbf{w} = 0$, which is a generalized eigenvalue problem and the eigenvector \mathbf{w}^* of $\mathbf{S}_W^{-1} \mathbf{S}_B$ is the desired optimal direction. Given the learnt LDA projection \mathbf{w}^* , a query face is classified into one of the two genders. The projection of feature vector \mathbf{x}_q (of a face with unknown gender) on the LDA space is given by $x_q^* = (\mathbf{w}^*)^\top x_q$.

Gender classification is performed based on the distance between the x_q^* and the means of the projected classes μ_1^* and μ_2^* such that

$$C_q = \begin{cases} 1 & \text{if } \|\mu_1^* - x_q^*\|_2 < \|\mu_2^* - x_q^*\|_2, \\ 2 & \text{otherwise} \end{cases}, \quad (8)$$

where $\mu^* = (\mathbf{w}^*)^\top \mu$

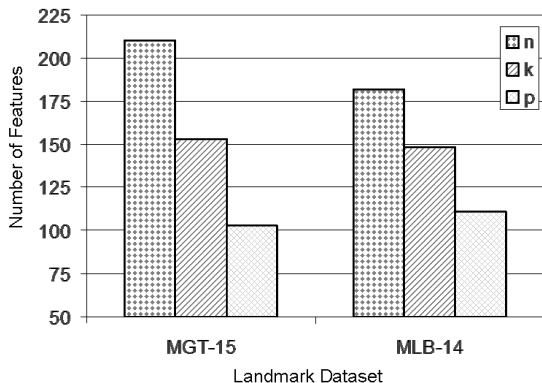


Fig. 6. Feature selection with 3D Euclidian and Geodesic distances using 15 MGT and 14 MLB landmarks. n denotes the original number of features, k denotes the candidate features while the number of compact feature set is denoted by p .

IV. EXPERIMENTATION AND RESULTS

We have performed gender classification on neutral as well as non-neutral 3D images of FRGC v2. In Experiment-1 we classify gender using all possible combinations of 3D Euclidean and geodesic distances between the 15 MGT manual [16] and 14 MLB [15] automatically detected landmarks. We also compare the classification accuracy using the 11 common landmarks. Refer to Figure. 2 for visual details of these landmarks. In Experiment-2, from the MLB set, we add the three missing landmarks to the set of 15 MGT landmarks bringing the total to 18. We then compare the classification accuracy using the 14 common landmarks on both datasets. Finally, in Experiment-3 we perform gender classification on the 466 images of individual subject in FRGC v2 dataset. 10 Fold cross validation technique is used in all experiments. Details on the experiments and their results are given in subsequent sections.

A. Experiment-1

We classify gender using 105 3D Euclidean and geodesic distances (total 210) obtained from the 15 manually annotated MGT landmarks and 91 3D Euclidean and geodesic distances (total 182) extracted from the 14 MLB automatically detected landmarks. For this purpose, we use the 4007 multiple expression scans of FRGC v2. We also perform gender classification using 55 3D Euclidean and geodesic distances (total 110) obtained from 11 common landmarks of both data sets (see Figure. 2 for landmark details). This is done to evaluate the validity of features extracted using automatically detected landmarks. Results are given in Table II. Figure. 7 (first two columns) show the ten most significant 3D Euclidean and geodesic features extracted using the 11 common landmarks.

B. Experiment-2

In Experiment-2 we add the missing Nasion (N), Subnasale (Sn) and Pogonion (Pg) from the set of automatically detected landmarks to the set of 15 MGT manually annotated ones and perform gender classification using 153 3D Euclidean and geodesic distances (total 306). Next, from the augmented 18 MGT landmark set we select the 14 landmarks that are common in both datasets and extract 91 3D Euclidean and

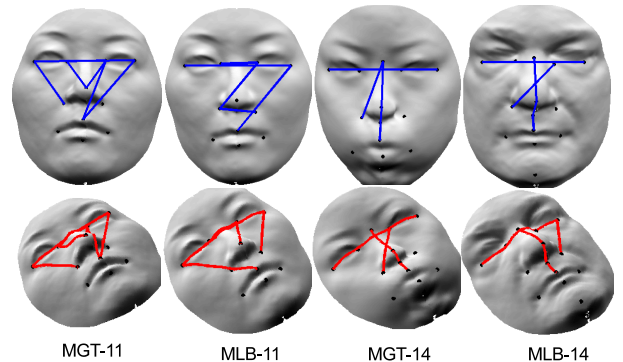


Fig. 7. Ten most significant features extracted from 11 common (first two columns) as well as the 14 common (last two columns) landmarks on MGT and MLB datasets. Notice the shift in significance of features extracted in MGT-14 when N, Sn and Pg are added to this set from MLB-14.

TABLE II. GENDER CLASSIFICATION ACCURACY RESULT (IN %) OF EXPERIMENTS 1 AND 2. BOLD FIGURES DENOTE BEST CLASSIFICATION ACCURACY IN EACH EXPERIMENT SUBCATEGORY.

	Experiment - 1				Experiment - 2			
	15	14	11	11	18	14	14	14
Number of landmarks used	15	14	11	11	18	14	14	14
Number of features extracted	105	91	55	55	153	91	91	91
Set of landmarks used	MGT	MLB	MGT	MLB	MGT	MLB	MGT	MLB
Accuracy with 3D Euclidean distances	85.71	87.48	84.15	82.11	88.15	87.48	87.83	87.48
Accuracy with geodesic distances	90.78	92.81	89.43	86.38	93.72	92.81	93.28	92.81
Accuracy with Combined Euclidean and geodesic	92.34	96.12	90.32	88.45	97.24	96.12	97.03	96.12

geodesic distances (total 182) to classify gender. This part of the experiment enables us to evaluate the classification significance of the three added landmarks. Comparative results are given in Table II. Ten most significant 3D Euclidean and geodesic features using the 14 common landmarks on MGT and MLB dataset are shown in Figure.7 (last two columns). All tests are carried out on the 4007 scans of FRGC v2.

C. Experiment-3

Here we classify gender on 466 scans of individual subjects of FRGC ver2 dataset using the features extracted from 14 MLB automatically detected landmarks. We then compare our classification accuracy with that of Ballihi et al. [9] and note that our selected features outperform the results of latter by about 11%. Comparative results are given in Table III.

TABLE III. COMPARATIVE GENDER CLASSIFICATION ACCURACY RESULT (IN %) FOR EXPERIMENTS 3.

Features	Ballihi et al. [9]	This work
	Radial and Circular geodesic curves	3D Euclidean and geodesic distances between biologically relevant landmarks
3D Euclidean	Not Reported	87.63
Geodesic	86.05	94.01
Combined	Not Reported	97.05

V. ANALYSIS

Results of Experiment-3 show that gender classification accuracy using geodesic distances extracted from biologically relevant landmarks is significantly higher (8%) than features extracted from arbitrary points on the face. It is hence evident

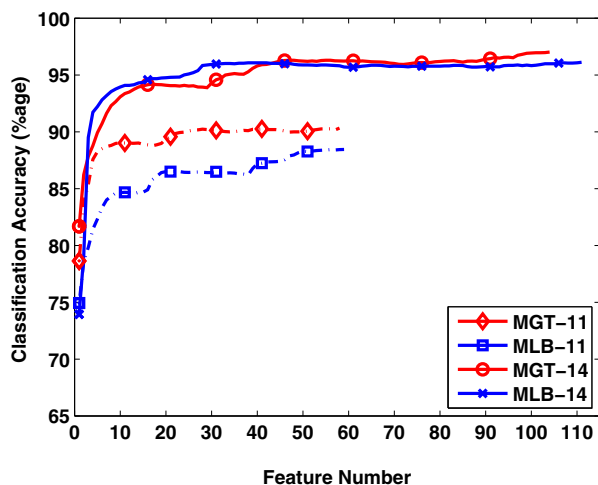


Fig. 8. Feature selection with 3D Euclidean and Geodesic distances using 11 and 14 common features on MGT and MLB Landmark datasets. Graph shows the final number of selected features and their respective classification accuracy.

that biological significance of features plays an important role in gender classification. Perhaps this is the main reason why our gender classification accuracy is much higher than that of Ballihi et al. [9]. Features used by Ballihi et al. do not model the strong gender differentiating parameters as compared to distances extracted between biologically significant anatomical landmarks. We also note that geodesics seem to be superior to 3D Euclidean distances in classifying gender. 3D Euclidean distances between biologically relevant landmarks still do not perform significantly better than the radial and circular geodesics extracted in [9] using arbitrary points. This is because of the fact that Euclidean distances cannot model the surface curvature between two given points (Figure. 1).

Even while using features related to biologically relevant points, it becomes clear from Experiments-2 and 3 that some features contain more gender differentiating information than others. The 15 MGT landmarks which are devoid of Nasion (N), Subnasale (Sn) and Pogonion (Pg) consistently perform below the accuracy level of 14 MLB landmarks which contain these points. The 14 MLB landmarks loose this supremacy in classifying gender when the three landmarks are added to the MGT set and it can be seen that the performance of both sets of landmarks is more or less the same. It is interesting to see how the inclusion of three biologically relevant gender differentiating landmarks can improve the classification accuracy by more than 6%. Notice from Figure.7 (last column) that nasal bridge length (N-Prn) and nasal height (Sn-Prn) stand out as one of the most significant features in classifying the gender. This result corroborates the findings of researchers [4], [5] who claim that the male nose is longer and more pointed than that of their female counterparts.

When comparing both sets of landmarks over distances extracted from 11 common points, MGT landmarks clearly outperform the MLB landmarks. This is due to some localization error in the landmarks of the latter. The same behaviour is repeated when Nasion (N), Subnasale (Sn) and Pogonion (Pg) are added to the common set of landmarks. However, now the difference in the superiority of MGT landmarks is reduced due to addition of three automatically detected landmarks which contain inherent localization error. Figure. 8 shows the classification accuracy using the combined 3D Euclidean and geodesic distances extracted from 11 and 14 common MGT and MLB landmarks.

We also perform an analysis of variation in the geodesic distances under different expressions. For this purpose we select all subjects with eight or more scans (1537 faces in total) in the FRGC ver2 dataset. Figure. 9 shows five distances that vary least in all faces and five that have the maximum variation. It is interesting to note that our algorithm gives good results on a database containing these variations.

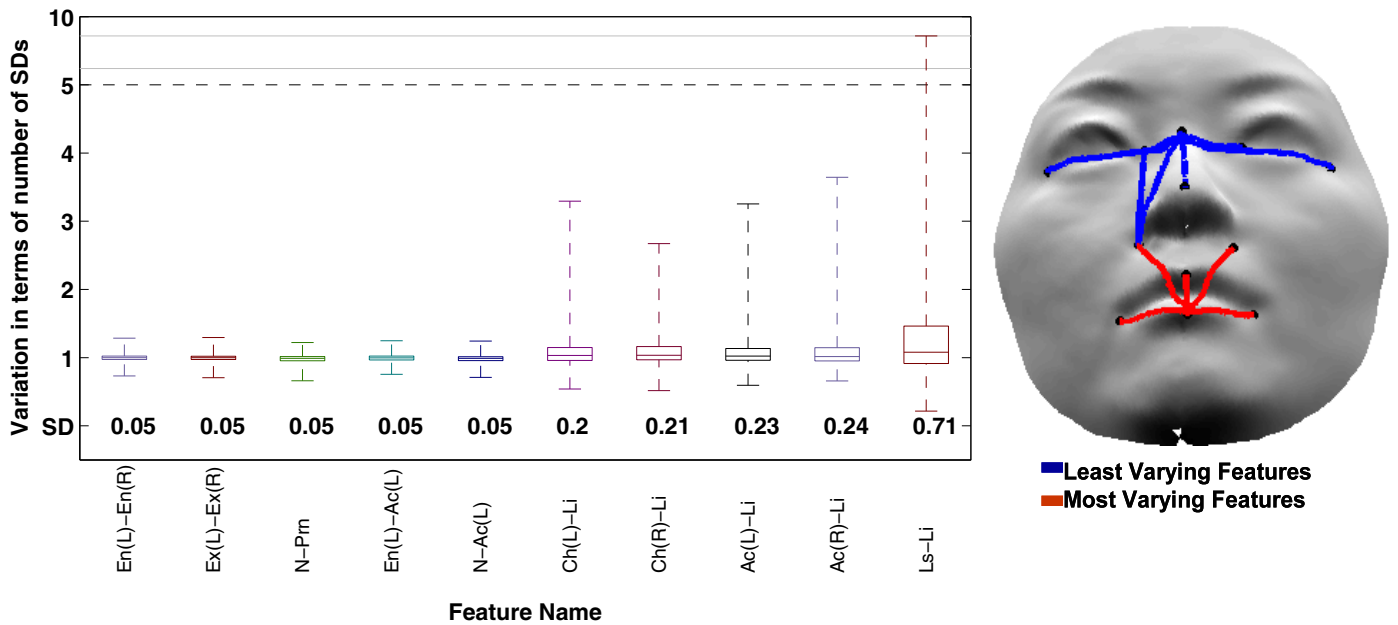


Fig. 9. Analysis of variation in geodesic distance under expression. Left: Five least and five most varying geodesics on 1537 scans of FRGC v2. Right: Graphical representation of these distances.

VI. CONCLUSION

In this paper we present a gender classification scheme that is robust to expression on a dataset that contains multiple ethnicities and pose variations. We show that a combination of 3D Euclidean and geodesic distance extracted from biologically relevant anatomical facial landmarks outperforms the classification accuracy obtained from arbitrary facial features. We report gender classification accuracy of 96.12% on 4007 scans of FRGC v2, one of the largest 3D face databases.

ACKNOWLEDGMENT

Syed Zulqarnain Gilani is funded by the International Postgraduate Research Scholarship (IPRS). This research was also supported by ARC grant DP110102399 and the UWA FECM grant.

REFERENCES

- [1] C. B. Ng, Y. H. Tay, and B. M. Goi, "Vision-based human gender recognition: A survey," *arXiv preprint arXiv:1204.1611*, 2012.
- [2] A. M. Burton, V. Bruce, N. Dench *et al.*, "What's the difference between men and women? evidence from facial measurement," *Perception*, vol. 22, pp. 153–176, 1993.
- [3] V. Bruce, H. D. Ellis, F. Gibling, and A. Young, "Parallel processing of the sex and familiarity of faces," *Canadian Journal of Psychology*, vol. 41, no. 4, p. 510, 1987.
- [4] D. H. Enlow, *Handbook of facial growth WB Saunders*. Philadelphia, 1982.
- [5] J. Shepherd, "The face and social attribution," *Handbook of Research on Face Processing*, pp. 289–320, 1989.
- [6] L. Farkas, *Anthropometry of the head and face, 2nd Ed.* Raven, 1994.
- [7] X. Han, H. Ugail, and I. Palmer, "Gender classification based on 3D face geometry features using SVM," in *CyberWorlds*. IEEE, 2009, pp. 114–118.
- [8] X. Lu, H. Chen, and A. Jain, "Multimodal facial gender and ethnicity identification," *Advances in Biometrics*, pp. 554–561, 2005.
- [9] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3D-geometric features for efficient face recognition and gender classification," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1766–1779, 2012.
- [10] Y. Hu, J. Yan, and P. Shi, "A fusion-based method for 3D facial gender classification," in *Computer and Automation Engineering (ICCAE)*, vol. 5. IEEE, 2010, pp. 369–372.
- [11] A. Hamza and H. Krim, "Geodesic matching of triangulated surfaces," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2249–2258, 2006.
- [12] S. Gupta, M. Markey, and A. Bovik, "Anthropometric 3D face recognition," *International Journal of Computer Vision*, vol. 90, no. 3, pp. 331–349, 2010.
- [13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [14] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, 2005, pp. 947–954.
- [15] C. Creusot, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3D meshes," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 146–179, 2013.
- [16] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *Biometrics: Theory, Applications, and Systems (BTAS)*. IEEE, 2009, pp. 1–6.
- [17] J. D'Erico, "Surface fitting using gridfit," in *MATLAB Central File Exchange*, 2008.
- [18] L. Farkas and I. Munro, *Anthropometric facial proportions in medicine*, 1987.
- [19] L. Farkas, "Anthropometry of the head and face in clinical practice," *Anthropometry of the Head and Face, 2nd Ed*, pp. 71–111, 1994.
- [20] K. Aldridge, I. George, K. Cole, J. Austin, T. Takahashi, Y. Duan, and J. Miles, "Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes," *Molecular Autism*, vol. 2, no. 1, p. 15, 2011.
- [21] I. M. Scott, N. Pound, I. D. Stephen, A. P. Clark, and I. S. Penton-Voak, "Does masculinity matter? the contribution of masculine face shape to

- male attractiveness in humans,” *PLoS one*, vol. 5, no. 10, p. e13585, 2010.
- [22] C. W. Bardin and J. F. Catterall, “Testosterone: A major determinant of extragenital sexual dimorphism,” *Science*, vol. 211, pp. 1285–1294, 1981.
- [23] D. H. Enlow, *Facial Growth (3rd ed.)*, 1990.
- [24] N. Pound, I. S. Penton-Voak, and A. K. Surridge, “Testosterone responses to competition in men are related to facial masculinity,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1654, pp. 153–159, 2009.
- [25] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial Intelligence*, vol. 97, no. 1, pp. 273–324, 1997.
- [26] R. Duda, P. Hart, and D. Stork, *Pattern classification and scene analysis 2nd ed.*, 2001.