Proc. of International Joint Conference on Biometrics (IJCB), (Washington DC, USA), October 2011.

Can Facial Metrology Predict Gender? *

Deng Cao, Cunjian Chen, Marco Piccirilli,
Donald Adjeroh, Thirimachos Bourlai, Arun Ross
West Virginia University, Morgantown WV 26506, USA
{dcao, cchen10, mpiccir1}@mix.wvu.edu
{Donald.Adjeroh, Thirimachos.Bourlai, Arun.Ross}@mail.wvu.edu

Abstract

We investigate the question of whether facial metrology can be exploited for reliable gender prediction. A new method based solely on metrological information from facial landmarks is developed. Here, metrological features are defined in terms of specially normalized angle and distance measures and computed based on given landmarks on facial images. The performance of the proposed metrology-based method is compared with that of a state-of-the-art appearance-based method for gender classification. Results are reported on two standard face databases, namely, MUCT and XM2VTS containing 276 and 295 images, respectively. The performance of the metrology-based approach was slightly lower than that of the appearance-based method by only about 3.8% for the MUCT database and about 5.7% for the XM2VTS database.

1. Introduction

Gender classification is a fundamental task for both humans and machines, as many social activities depend on precise gender identification. The problem has attracted considerable attention, and has been investigated from both psychological [5, 9] and computational [7] perspectives. Various studies have attempted to perform gender classification using human gait [14] or body information [6]. Introduced in the 1990s, SEXNET was among the first automated systems capable of performing gender identification using human faces [11]. Since then, a number of studies have investigated the problem as part of the general face recognition (FR) problem. Modern FR systems typically combine textural information from the face with facial geometry. Popular examples include active appearance models [13], local feature analysis [12], and elastic bunch graph matching [28]. In such systems, the information about facial geometry is often captured from specific landmarks on the face. In this work, we investigate whether topological information extracted from facial landmarks can be used to efficiently perform gender classification.

Facial landmarks can be divided into three broad categories [24]: anatomical landmarks, mathematical landmarks, and pseudo-landmarks. Anatomical landmarks are biologically meaningful points defined as standard reference points on the face and head, such as pupil, dacryon, nasion or pogonion, and they are often used by scientists and physicians. They tend to be somewhat more abstract than other features of the skull (such as protuberances or lines). Anatomical landmarks are considered very important because they are useful in various scientific fields including computer vision, cosmetic surgery, anthropology, and forensics. However, because of various types of distortions, it is quite difficult to accurately extract anatomical landmarks from 2D face images, either manually or automatically. Furthermore, the number of useful anatomical landmarks that can be extracted from a single face image is considered limited. Thus, the exclusive use of anatomical landmarks in face recognition is not recommended.

Mathematical landmarks are defined according to certain mathematical or geometric properties of human faces, such as middle point between two anatomical landmarks, extreme point with respect to particular face region (for example, leftmost point of face contour), or centroid of a certain group of landmarks. A mathematical landmark may or may not coincide with an anatomical landmark, and it can be easily located using automated methods. Facial pseudo-landmarks are defined based on two or more mathematical or anatomical landmarks, or around the outline of facial or hair contours. They are not rigorously defined, and can be approximately located using prior knowledge of anatomical or mathematical properties. Pseudo-landmarks are relatively easy to acquire, and are generally accurate enough for appearance-based face recognition methods. In this work, pseudo-landmarks are used to extract facial metrology.

*This project was funded by the Center for Identification Technology Research (CITeR).

Proc. of International Joint Conference on Biometrics (IJCB), (Washington DC, USA), October 2011.

1.1. Benefits of Facial Metrology

In traditional face recognition, facial landmarks can be either automatically detected or manually annotated. These landmarks are often used for registration purposes [4] and they can assist the recognition system mainly as part of the pre-processing step. There are many advantages to the use of facial metrology. These include (i) *Memory Management*: compared to texture-based information in face images, landmarks require much less storage space (each landmark consists of two numbers, i.e., the row and column on a 2D face canonical coordinate system); (ii) *Information Privacy*: unlike the full face image, landmark information can be safely stored, transported, and distributed without potential violation of human privacy and confidentiality; (iii) *Prediction of Missing Information*: topological features (face coordinates) can be either global or local to specific facial regions. Thus, missing information can be approximately predicted, for example, using statistical approaches [1]; (iv) *Law Enforcement*: useful information from facial metrology could be used as forensic evidence in a court of law, where admissibility of quantifiable evidence is a major consideration.

1.2. Main Contribution

The main challenges related to facial metrology include (a) precision in localizing the landmark coordinates; (b) sensitivity of landmark localization to pose, expression, and aging; and (c) sparsity of information encoded in landmarks for human identification. In spite of these challenges, we believe that landmarks from 2D faces can provide important cues for problems related to human recognition. Following a recent study on predictability and correlation in whole-body human metrology [1], in this paper we hypothesize that the information extracted from facial metrology exclusively can be used for gender recognition or facial classification. The main goal of this work is to determine whether facial metrology can be used to discriminate between genders. We assume that facial landmarks on a given face image are already known and our research focus is to perform gender classification based solely on the information provided by facial landmarks. If gender classification can be successfully performed using these landmark points, then investment can be made in automating the landmark detection process. The performance of our proposed facial metrology-based gender classification algorithm is compared to benchmark appearance-based technique, viz., the Local Binary Patterns (LBP) method. The main contribution of our work is a demonstration that the classification performance of our method is comparable to that of an appearance-based method. In practice, we illustrate that by using only weak features, i.e., facial metrological features derived from landmarks, our approach results in only about 3.8-5.7% lower classification rate (on two different face databases) compared to a benchmark appearance-based method.

2. Related Work

Humans perceive gender not only based on the face, but also on the surrounding context such as hair, clothing and skin tone [15, 6], gait [14] and the whole body [6, 1]. Below, we review relevant work on gender prediction from facial images only.

The problem of gender classification based on human faces has been extensively studied in the literature [20, 3]. There are two popular methods. The first one is proposed by Moghaddam et al. [20] where a Support Vector Machine (SVM) is utilized for gender classification based on thumbnail face images. The second was presented by Baluja et al. [3] who applied the Adaboost algorithm for gender prediction. Recently, due to the popularity of Local Binary Patterns (LBP) in face recognition applications [2], Yang et al. [30] used LBP histogram features for gender feature representation, and the Adaboost algorithm to learn the best local features for classification. Experiments were performed to predict age, gender and ethnicity from face images. A similar approach was proposed in [25]. Other local descriptors have also been adopted for gender classification. Wang et al. [27] proposed a novel gender recognition method using Scale Invariant Feature Transform (SIFT) descriptors and shape contexts. Once again, Adaboost was used to select features from face images and form a strong classifier.

Gao et al. [10] performed face-based gender classification on consumer images acquired from a multi-ethnic face database. To overcome the non-uniformity of pose, expression, and illumination changes, they proposed the usage of Active Shape Models (ASM) to normalize facial texture. The work concluded that the consideration of ethnic factors can help improve gender classification accuracy in a multiethnic environment. A systematic overview on the topic of gender classification from face images can be found in [17]. Among all the descriptors that encode gender information such as LBP [25], SIFT [26] and HOG [6], the LBP has shown good discrimination capability while maintaining simplicity [17]. To establish a base-line for appearance-based methods, we use LBP in combination with SVM to predict gender from facial images in this work.

Although in previous work [22, 29] geometry features were used as *a priori* knowledge to help improve classification performance, none of the aforementioned approaches, unlike our work, focused explicitly and solely on facial metrology as a means for gender classification. Perhaps our work is more closely related to earlier research by Shi et al. [23, 24] on face recognition using geometric features, where they used ratio features computed from a few anatomical landmarks. However, we take a more comprehensive look at the explicit use of facial geometry in solving the problem

Proc. of International Joint Conference on Biometrics (IJCB), (Washington DC, USA), October 2011.

of gender classification. We use solely metrological information based on landmarks, which may or may not be biologically meaningful. In our approach, the local information from independent landmarks is used instead of holistic information from all landmarks.

3. Gender Prediction via Facial Metrology

3.1. Facial Landmarks

Two well-known databases were used in this work, namely, MUCT [19] and XM2VTS [18]. For each subject in each database, only one frontal face image and the corresponding landmark information were used. Compared to the XM2VTS database, the MUCT database has more diversity with respect to facial expressions, pose, and ethnicity. In particular, MUCT has much more variation in mouth shapes. Figure 1 shows two sample faces with numbered landmarks, one from each database. The numbering system used in XM2VTS is the same as that of MUCT, except for a set of extra landmarks used in MUCT (i.e., #69 - #76). Details about the databases can be found in the section on experiments.

Before extracting measurements from the face, we first consider the spatial distribution of facial landmarks in the faces in the databases. Such a distribution could shed some light on the potential of landmarks in gender prediction. Let N be the number of landmarks for each face. A face, F^k, can be represented as a vector

$$F^k = (x_1^k, y_1^k), (x_1^k, y_1^k), \ldots, (x_i^k, y_i^k), \ldots, (x_N^k, y_N^k) \quad (1)$$

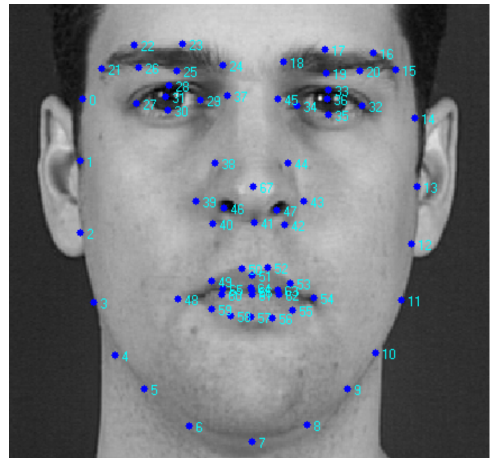
where (x_i^k, y_i^k) is the Cartesian coordinate of the i-th landmark of F^k, $k = 1, 2, \ldots, m$, and m is the number of faces in the training set. To compute the landmark distribution, each landmark $L_i^k = (x_i^k, y_i^k)$ is normalized as follows:

$$\hat{x}_i^k = \mu(x_i) + \alpha \left(\frac{x_i^k - \mu(x_i)}{\sigma(x_i)} \right) \quad (2)$$

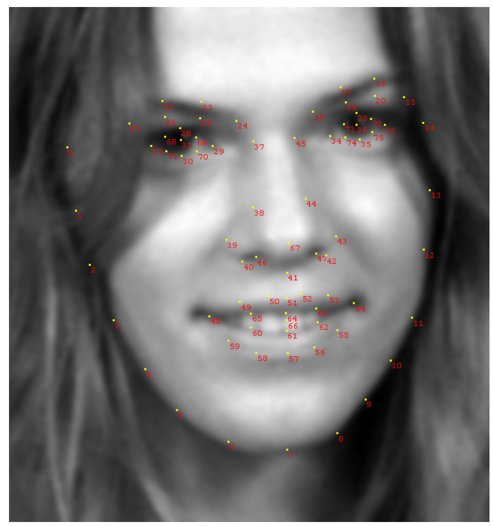
$$\hat{y}_i^k = \mu(y_i) + \alpha \left(\frac{y_i^k - \mu(y_i)}{\sigma(y_i)} \right) \quad (3)$$

where α is a constant, and $\mu()$ and $\sigma()$ are the mean and standard deviation of each landmark location, respectively.

Figure 2 shows the landmark distribution for male and female subjects in the two databases. The landmark coordinates are aligned by the nose tip position, which is landmark #67 in both databases. As the figure shows, some landmarks show a significant separation between male and female subjects, while others are difficult to separate. Normalized landmark positions with more separation between the male and female subjects are likely to result in better gender classification performance.



(a)



(b)

Figure 1. Sample faces with numbered landmarks from (a) XM2VTS and (b) MUCT.

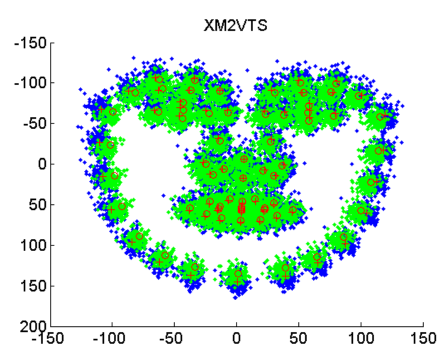


Figure 2. Landmark distribution for the XM2VTS database using $\alpha = 5$. The red cross and red circle indicate average landmark positions across male and female subjects, respectively. The blue and green scatter points are normalized landmark coordinates for each individual (blue for males, green for females).

3.2. Metrological Features

There are different ways to utilize the landmark information. We cannot directly use the landmark coordinates, since they will be sensitive to translation, scaling, and 2D rotation of face images. One could consider all distance ratios defined by sets of four landmarks, or triangular features defined by any three non-collinear landmarks [24]. The problem here is computational complexity. The dimensionality of the feature space will be $\Theta(N^4)$ for complete distance ratios, and $\Theta(N^3)$ for landmark triplets. An alternative is to consider simple Minkowski distances between two arbitrary landmarks $L_i^k = (x_i^k, y_i^k)$ and $L_j^k = (x_j^k, y_j^k)$, given by:

$$D_{ij}^k = \left((x_i^k - x_j^k)^p + (y_i^k - y_j^k)^p\right)^{\frac{1}{p}}, \tag{4}$$

where p is the distance parameter. For a given p, the number of distances is thus $\Theta(N^2)$. In this work we have considered the Euclidean distance ($p = 2$). The distances can be easily normalized to be scale-invariant by a reliable measure, such as inter-eye distance. The resulting ratios are also invariant to translation and 2D rotation. However, using only distance measures may not be reliable, since the orientation of the distance vectors may be significant as well. For example, two individuals may have the same distance from the tip of the nose to the pupil, although one may have a longer face and the other may have more widely separated eyes. To improve the reliability of the features, we also use the horizontal angle subtended by each distance vector. The horizontal angle A_{ij}^k is computed from the pair-wise landmark coordinates:

$$A_{ij}^k = arctan\left(\frac{y_i^k - y_j^k}{x_i^k - x_j^k}\right) \tag{5}$$

3.3. Feature Ranking and Selection

Using all pair-wise distances will lead to a very high dimensional feature space. For example, there are 5,700 features (distances and angles) for the MUCT database. Another problem is that the distance and angle measures may not always be robust. Some features may not be useful for gender discrimination and some others may be sensitive to errors caused by inconsistent landmark positions. Performance will be compromised if such features are not removed or their impact minimized.

To handle these issues, we apply a simple, yet efficient and robust, d-prime-like scheme to rank the distances by their discrimination capabilities. For each pair-wise distance, and across all the faces in our training set (see definition in experiments section), we compute the d-prime as follows:

$$d'_{ij} = \frac{\mu(D_{ij}^M) - \mu(D_{ij}^F)}{\sqrt{([\sigma(D_{ij}^M)]^2 + [\sigma(D_{ij}^F)]^2)/2}} \tag{6}$$

where $\left(\mu(D_{ij}^M), \mu(D_{ij}^F)\right)$ and $\left(\sigma(D_{ij}^M), \sigma(D_{ij}^F)\right)$ are the mean values and variances of the distance distributions between landmark i and j, respectively, and M and F denote male and female, respectively. Similarly, we compute the d-prime-like value for each angle. If the two distributions are well separated, the d-prime value should be relatively high. Otherwise, the measure results in a high intra-class error and should not be considered as a reliable feature. The measures are then ranked in decreasing order based on their d-prime values, which corresponds to a decreasing order in their gender discrimination ability. Figure 3 shows a sample face annotated with the top-20 ranked landmarks. Our results show that only a few top-ranked features are necessary for gender discrimination purposes. Note that the specific d-prime ranking for a given measurement could vary from database to database. However, the general trend is similar when both datasets were used (see also Table 1 in Section 5.3).

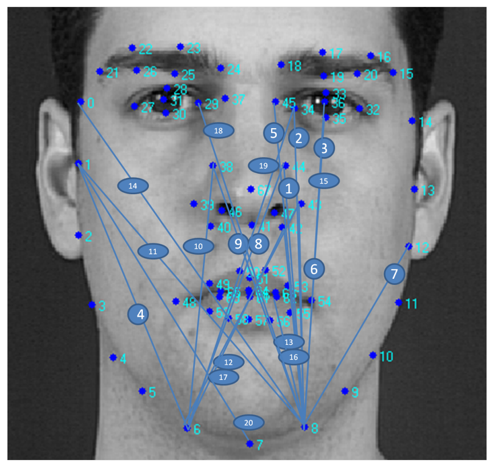


Figure 3. The top 20 pairwise distances ranked by their discrimination ability (Eq. 6). The sample face is from the XM2VTS database. Numbers on the edges indicate the d-prime ranking.

3.4. SVM Classifier

We tested and compared results using three classifiers: support vector machine (SVM), k-nearest neighbor (KNN) and logistic models. We chose SVM for its superior performance and speed. For the SVM, we used the Gaussian radial basis function as the kernel:

$$K(u, v) = exp(-\frac{||u - v||^2}{2\sigma^2}) \tag{7}$$

where u and v are the feature vectors, and σ is the width of the basis function. We set σ=2. The SVM soft margin parameter C [8]is set to be 10.

The experimental results suggest that among the thousands of metrology-based features, only a few top-ranked

Proc. of International Joint Conference on Biometrics (IJCB), (Washington DC, USA), October 2011.

features are significant in gender prediction. However, the optimal number of features depends on the database used. More measures may not necessarily improve the performance. Instead, too many measures may introduce more noise thereby compromising the performance (see Figure 6 under experimental results).

4. Gender Prediction via Appearance

To compare our results with state-of-the-art approaches, the use of appearance-based models for gender recognition was considered. In particular, we considered the use of the LBP operator on the same datasets as we used for the study of metrological features. The basic LBP descriptor encodes micro-patterns of an image by thresholding 3×3 neighborhoods based on the value of the center pixel and then transforming the converted binary pattern sequence into a decimal value. It can be extended to accommodate neighborhoods of different sizes to capture textures at multiple scales [21].

To utilize LBP method for the extraction of gender features from facial images, the input image is first divided into non-overlapping blocks. Then, the spatial histogram features from each block is calculated and concatenated to form a global descriptor. Here, the LBP operator is denoted as $LBP^{u^2}_{P,R}$, where P refers to the number of equally spaced points placed on a circle with radius R and u^2 represents the uniform concept, which accounts for most of the patterns observed in the experiment. For instance, 11001111 is considered to be a uniform pattern since it contains no more than 2-bitwise transition (1 to 0 and 0 to 1). When computing LBP histograms, every uniform pattern has a separate bin (58 bins in total) and all the other non-uniform patterns together have a single bin. In our experiments, the $LBP^{u^2}_{8,2}$ descriptor is used. The image is resized to 126×90, with each block consisting of 18×15 pixels. The total number of blocks is, therefore, $7 \times 6 = 42$. For each block, we use LBP to extract 59 bin features, leading to a 2478-dimensional feature vector (see Figure 4).

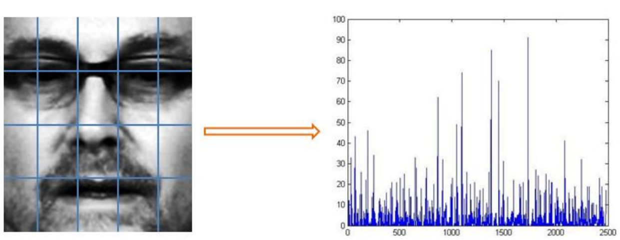


Figure 4. LBP gender feature representation. The face image is acquired from MUCT database [19].

In order to design the gender classifier (i.e., predictor), SVM is used. The SVM classifier is trained using a training set of labeled face images. The test sample is classified according to the sign of $y(s)$,

$$y(s) = w^T \phi(s) + b, \tag{8}$$

where $\phi(s)$ denotes the transformation of the original feature-space and b is the bias. w is the normal vector and determines the orientation of the hyper plane which is generated during SVM training. For classification, we use the histogram intersection kernel:

$$k(x, y) = \sum_{i=1}^{n} min(x_i, y_i), \tag{9}$$

where x_i and y_i are the i^{th} histogram bin for the feature vectors of x and y. The histogram intersection kernel was observed to be much more effective for classification than the linear or the RBF kernel when the LBP histogram sequence features were used as input. Therefore, it is adopted in our appearance-based gender classification scheme.

5. Experiments

5.1. Datasets and Setup

The MUCT land-marked face database [19] was created by researchers to generate data exhibiting diversity in pose, illumination, age, and ethnicity. We use 276 subjects from Category-A, consisting of 131 males and 145 females. The first sample of each subject (usually the frontal face) is selected and used in our experiments. Therefore, a total of 276 samples are used. The size of the original image is 480×640 pixels. Since the landmark positions for the eyes are provided, for the appearance-based LBP method, the images are normalized and aligned based on eye coordinates. Further, for the LBP method, all samples are preprocessed with histogram equalization to reduce the effect of illumination. The final cropped image size is set to be 130×150. For the LBP method, it would be resized to 126×90. Sample images are shown in Figure 5.

The XM2VTS database [18] has 295 subjects. Each subject has one sample selected. There are 160 males and 135 females. Similar to the MUCT samples, the size of the cropped sample images is also 130×150.

To perform gender classification on the MUCT database, we randomly select 50 males and 50 females for training. The remaining 176 samples are reserved for testing. This partitioning exercise is repeated 50 times. For XM2VTS, the same experimental design was applied, except the total number of test samples in this case was 195.

For the metrology-based approach, the d-prime-like feature ranking is applied separately on both the distance measures and the angle measures. Thus, we use both top-ranked distances and top-ranked angles for the analysis.

Proc. of International Joint Conference on Biometrics (IJCB), (Washington DC, USA), October 2011.

Figure 5. Sample images from the MUCT database. Face images in the bottom row correspond to the cropped and geometrically normalized images, after face detection on the top row images.

5.2. Results

Figure 6 shows the performance of the metrology-based method for gender classification, using the proposed metrology-based features from the facial landmarks. The performance using distances and angles separately varied somewhat with the database. In most cases, the performance on MUCT database is slightly better than that on XM2VTS. Further, the distance measures performed generally better than the angle measures. However, fusing the distance and angle measures at the feature level generally improved classification performance, especially when the feature space is small (less than 80 features). We observe that the metrology-based system can provide good results with only a few landmarks (around 10), implying that there is a possibility of using a lower-dimensional space, and hence lower computational cost. However, the experimental results do not indicate whether there exists an optimal number or combination of features. Since a large feature space will not necessarily lead to superior performance, we selected only the top-10 ranked distances and the top-10 ranked angles for subsequent experiments [1].

5.3. Landmark Discrimination Ability

The discussion so far has focused on the distance between pairs of landmarks, or the angles formed by such distance vectors. We also evaluated the discrimination capability of the *individual* landmarks. While we expect a landmark with high discrimination ability to be involved in distance or angle measures with an equally high discrimination ability, this may not always be the case. An evaluation of the discrimination ability of individual landmarks is important in identifying landmarks that are major determinants of performance in a metrology-based method. Such

[1]Due to the large number of features present in this work, we did not apply any standard feature selection schemes for determining the optimal set of features. Such an experiment will be conducted in the future.

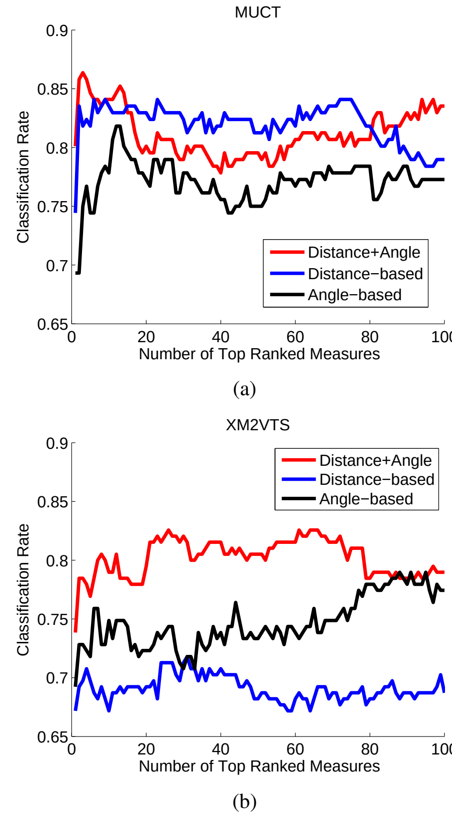


Figure 6. Performance comparison using the Top 1-100 angles, Top 1-100 distances and their fusion (2-200 features) in gender classification.

landmarks can then become the focus of a more concerted effort at automated landmark detection. Consider the distances between landmark pairs as a matrix. To determine the discrimination ability of a single landmark, we simply compute the average d-prime value between that landmark and all the other landmarks. For the i-th landmark L_i, we have:

$$d_i^{'} = \frac{1}{N-1} \sum_{j \neq i} d_{ij}^{'}. \tag{10}$$

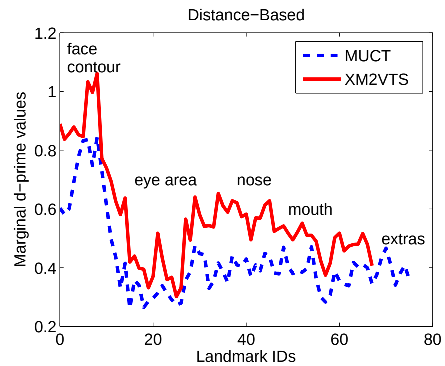
$d_i^{'}$ is called the marginal d-prime. We calculate the marginal d-prime values for distances and for angles separately.

Table 1 shows the top-20 landmarks, ranked by their discrimination capability, as determined by their marginal d-prime values. Figure 7 shows an annotated view of the discrimination capability of the landmarks and their approximate regions on the face. Our results indicate that from a distance-based perspective, the landmarks on the face contour are crucial for gender classification. This is not entirely surprising, especially given the original distribution of the landmarks on the face, as was shown in Figure 2. The landmarks in the eye region tend to have a low discrimination capability, perhaps due to the effect of the eyelids. The

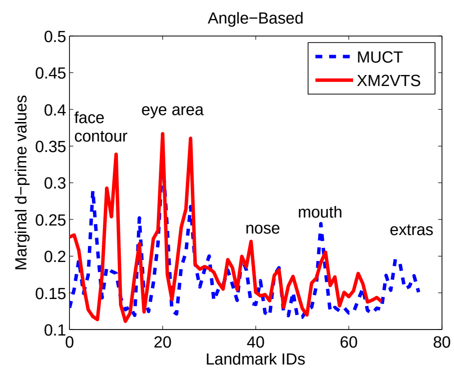
Table 1. Top 20 landmarks ranked with respect to their discrimination ability using distance (D) and angle (A) measures.

	MUCT		XM2VTS			MUCT		XM2VTS	
Rank	D	A	D	A	Rank	D	A	D	A
1	9	21	9	21	11	2	38	11	19
2	7	6	7	27	12	12	71	12	40
3	6	27	8	11	13	30	3	35	16
4	5	16	1	9	14	55	72	30	3
5	8	55	4	26	15	49	28	15	56
6	10	22	3	10	16	71	9	38	38
7	4	20	5	25	17	45	25	46	35
8	11	7	6	20	18	31	46	13	55
9	1	26	2	2	19	32	30	39	28
10	3	31	10	1	20	38	35	45	30

discrimination capability of landmarks in the nose region varied with different databases, probably due to the inconsistency in the annotation process. The landmarks in the mouth region also showed a low discrimination capability, because their positions are sensitive to the significant variability due to mouth expression. The angle-based marginal d-prime values are generally low, but they can still help in improving the gender recognition performance, as was shown in Figure 6. For the angular measurements, the landmarks on the face contour and in the eye region seem to be more significant than landmarks in other facial regions for the problem of facial metrology-based gender prediction.



(a)



(b)

Figure 7. Discrimination ability of individual landmarks (based on marginal d-prime values), along with the approximate facial regions for the landmarks. (a) distance-based; (b) angle-based.

5.4. Comparative Performance

The experimental results show that facial metrology do have the potential to discriminate between genders. To place the results of the facial metrology-based approach in context, we compared it with the results obtained using an appearance-based method for gender identification.

As shown in Figure 8 and Table 2, the current performance of the metrology-based approach is slightly lower than the appearance-based method. The major cause might be due to the limited nature of the information encoded in the landmarks, and the nontrivial human errors in the annotation process. Unlike in the classification of facial expression [16], we do not have prior knowledge about what local facial regions are most critical in determining gender. Yet, the performance of the metrology-based approach (86.83%, 82.83%) was only slightly inferior than that of the appearance-based method (90.63%, 88.56%) by about 3.8% for the MUCT database, and about 5.7% for the XM2VTS database. Also, compared to a 2478-dimensional feature space in LBP, the metrology-based method has a 20-dimensional feature space. Thus the execution time in the test stage of our metrology-based method is lower than that of the LBP method: 0.02 ms vs 1.8 ms per image for MUCT database, and 0.03 ms vs 1.7 ms for XM2VTS database.

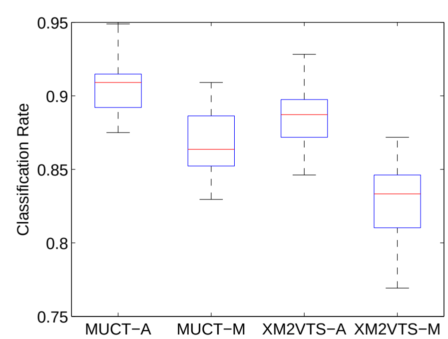


Figure 8. Comparative Performance: A–Appearance-based; M–Metrology-based.

6. Discussion and Conclusion

The results show that facial metrology can indeed be used for gender classification. There are still several interesting open questions that need to be further studied. A key question would be how to improve the performance of the metrology-based method. How can the metrology-based method maintain its performance when confronted with in-

Table 2. Summary of the comparative performance results when using facial metrology(top 10 landmarks) and appearance-based models in gender classification. The summary statistics in this table are associated with the results in Figure 8.

	Metrology-Based		Appearance-Based	
Classification Rate	MUCT	XM2VTS	MUCT	XM2VTS
Mean	0.8683	0.8283	0.9063	0.8856
Max	0.9091	0.8718	0.9489	0.9282
Min	0.8295	0.7692	0.8750	0.8462
Std	0.0217	0.0251	0.0168	0.0191

creasing database sizes, and more variabilities in the face, say due to pose, expression, race, aging, etc? The above two questions might be properly addressed by introducing a robust landmark detection technique, which can consistently localize the position of the landmarks. Another question has to do with the determination of the true capacity of facial metrology. The advantage of our proposed approach is that, due to its simplicity and independence, it could be combined with other more accurate (yet more computationally expensive) methods to improve the overall recognition performance. This paper is a good starting point in addressing these questions, especially for gender classification, and perhaps for the more general problem of face recognition.

References

[1] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross. Predictability and correlation in human metrology. *IEEE WIFS*, 2010.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE TPAMI*, 28:2037–2041, December 2006.

[3] S. Baluja and H. A. Rowley. Boosting sex identification performance. *IJCV*, 71(1):111–119, 2007.

[4] G. M. Beumer, Q. Tao, A. M. Bazen, and R. N. J. Veldhuis. A landmark paper in face recognition. In *FGR*, pages 73–78, 2006.

[5] 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.

[6] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang. Gender recognition from body. In *ACM Multimedia*, 2008.

[7] M. Castrillón-Santana and Q. C. Vuong. An analysis of automatic gender classification. In *CIARP*, 2007.

[8] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[9] B. Edelman, D. Valentin, and H. Abdi. Sex classification of face areas: how well can a linear neural network predict human performance. *Biological System*, 6(3):241–264, 1998.

[10] W. Gao and H. Ai. Face gender classification on consumer images in a multiethnic environment. In *ICB*, pages 169–178, 2009.

[11] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *NIPS*, pages 572–577, 1990.

[12] S. G. Kong, J. Heo, B. R. Abidi, J. K. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition - a review. *CVIU*, 97:103–135, 2005.

[13] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic face identification system using flexible appearance models. *IVC*, 13:393–401, 1995.

[14] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. *IEEE TSMC, Part C*, 38:145–155, 2008.

[15] X.-C. Lian and B.-L. Lu. Gender classification by combining facial and hair information. In *ICONIP (2)*, pages 647–654, 2008.

[16] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE TPAMI*, 21:1357–1362, 1999.

[17] E. Makinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE TPAMI*, 30:541–547, 2008.

[18] K. Messer, J. Matas, J. Kittler, and K. Jonsson. XM2VTSDB: The extended M2VTS database. In *AVBPA*, pages 72–77, 1999.

[19] S. Milborrow, J. Morkel, and F. Nicolls. The MUCT landmarked face database. *Pattern Recognition Association of South Africa*, 2010.

[20] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *IEEE TPAMI*, 24:707–711, 2002.

[21] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE TPAMI*, 24:971–987, 2002.

[22] Y. Saatci and C. Town. Cascaded classification of gender and facial expression using active appearance models. In *FGR*, 2006.

[23] J. Shi, A. Samal, and D. Marx. Face recognition using landmark-based bidimensional regression. In *ICDM*, pages 765–768, 2005.

[24] J. Shi, A. Samal, and D. Marx. How effective are landmarks and their geometry for face recognition? *CVIU*, 102(2):117–133, May 2006.

[25] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao. Gender classification based on boosting local binary pattern. In *ISNN (2)*, pages 194–201, 2006.

[26] M. Toews and T. Arbel. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE TPAMI*, 31(9):1567–1581, 2009.

[27] J. Wang, J. Li, W. Yau, and E. Sung. Boosting dense sift descriptors and shape contexts of face images for gender recognition. In *CVPRW*, pages 96–102, 2010.

[28] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE TPAMI*, 19:775–779, 1997.

[29] Z. Xu, L. Lu, and P. Shi. A hybrid approach for gender classification from face images. In *ICPR*, 2008.

[30] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *ICB*, pages 464–473, 2007.